



Translation of DNA into Evolvable Sequence-Defined Synthetic Polymers

Citation

Niu, Jia. 2014. Translation of DNA into Evolvable Sequence-Defined Synthetic Polymers. Doctoral dissertation, Harvard University.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:12274203>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

©2014 by Jia Niu

All rights reserved

Translation of DNA into Evolvable Sequence-Defined Synthetic Polymers

Abstract

Laboratory directed evolution have enabled the discovery of numerous functional natural and synthetic macromolecules with tailor-made functions. However, approaches that use enzymes to effect the crucial translation from an information carrier molecule such as DNA or RNA to synthetic polymers are limited to producing close analogs of nucleic acids, either due to a strict requirement to hybridize with a nucleic acid template or as a consequence of the limited substrate scope of polymerase enzymes.

In Chapter Two, we developed a DNA-templated translation system that enables the enzyme-free translation of DNA templates into sequence-defined synthetic polymers with no necessary structural relationship with nucleic acids. We demonstrate the efficiency, sequence-specificity, and generality of this translation system by polymerizing building blocks including polyethylene glycol (PEG), β -amino acids, and α -(D)-amino acids in a DNA-programmed manner. Sequence-defined synthetic polymers with molecular weights of up to 26 kDa containing 16 consecutively coupled building blocks can be translated from DNA templates using this strategy.

In Chapter Three, we further incorporated the translation strategy developed in Chapter Two into a synthetic polymer *in vitro* selection system that integrates a carefully

designed genetic code to enable a DNA template library with diversity of 7×10^{10} members, 16 macrocyclic substrates with structurally and functionally diverse β -peptide building blocks, optimized DNA-templated translation conditions, and molecular biology methods to prepare DNA templates surviving *in vitro* selection and PCR amplification for subsequent rounds of translation and selection. We anticipate these developments have the potential to enable the laboratory evolution of a wide range of synthetic polymers and to reveal the evolutionary potential of macromolecules beyond the reach of previous translation systems.

In Chapter Four, we developed a strategy to use T4 DNA ligase to mediate the ligation of short oligonucleotides with various different non-natural side-chain modifications. We described the incorporation of eight different functional groups encoded by eight trinucleotide codons. The translation system exhibits high sequence specificity and efficiency and could generate a polymer of 50 consecutive building blocks. Finally, we demonstrate an *in vitro* selection system of highly functionalized nucleic acid polymers based on the ligase-mediated translation strategy.

Table of Contents

Chapter One.....1

Introduction: Discovering Functional Sequence-Defined Synthetic Polymers by Directed Evolution

Chapter Two.....33

Enzyme-Free Translation of DNA into Sequence-Defined Non-Nucleic Acid Synthetic Polymers

Chapter Three.....88

Developments towards an *In Vitro* Selection System of Functional Synthetic Polymers

Chapter Four.....182

DNA Ligase-Mediated Translation of DNA into Highly Functionalized Nucleic Acid Synthetic Polymers

Acknowledgements

First and foremost I would like to thank my thesis advisor Professor David R. Liu for his teaching in science and in many other aspects of my scientific maturation. He has been a role model for me from my early scientific career even before we met in person, and I feel extremely lucky to be able to work with him in Harvard. I am also grateful for the research environment that he has created that leads to the successful conclusion of my graduate school.

I have been very fortunate to be able to work with the top-notch scientists during my PhD. Dr. Jennifer Heemstra was an early mentor of mine in graduate school, and she showed me the depth of nucleic acid chemistry which I resolved to further study for the rest of my scientific career. Dr. Yevgeny Brudno has been a tremendous source of inspiration and encouragement. His brilliance and enthusiasm on the DNA-templated polymerization project deeply influenced and inspired me. Dr. Ryan Hili is a great colleague that I fortunately to work with, we have achieved some very exciting results together that could have only come into my dream if I had not worked with him. I would also like to thank Zhen Chen who has been a very smart and diligent coworker in the synthetic polymer project, and he has done indispensable contribution to it.

My scientific development was also supported by many previous and current Liu group members, who have generously provide me with helps in science and in life. I would not be able to name all who have helped me, but still like to single out Dr. Yu He, Dr. Yiyun Chen, Dr. Ralph Kleiner, Dr. Yinghua Shen, Dr. Christoph Dumelin, Dr. Lynn

McGregor, Dr. John Guilinger, Dr. Richard McDonald, Dr. Bryan Dickinson, Dr. Xue Gao, Dr. Ning Sun, Dr. Dmitry Usanov, Chris Green, Dennis Dobrovolsky, and Juan Pablo Maianti.

I have been very fortunate to have the instruction from my committee members Prof. Jack Szostak and Prof. Alan Saghatelian. I very much enjoy hearing their guidance for my project and for greater scientific pursuit.

None of my work would become possible without the tremendous support from my parents, Zhimin Niu and Fengying Tan. They have allowed me to pursue my scientific dream from the very beginning, and are always there to support me whenever I need.

Last, I want to take this opportunity to thank my wife, Yanyan Lu. Her love and encouragement accompanied me since we were together. I have cherished her companion as the happiest thing ever happened to me.

Chapter One

**Introduction: Discovering Functional Sequence-Defined
Synthetic Polymers by Directed Evolution**

1.1 Biopolymers Versus Synthetic Polymers: What is Different?

The definition of polymer by International Union of Pure and Applied Chemistry (IUPAC) states that polymer is “*a molecule of high relative mass, the structure of which essentially comprises of the multiple repetition of units derived, actually or conceptually, from molecules of low relative molecular mass*”.¹ Such definition includes both synthetic polymers and biopolymers in the category of polymers. However, the current biopolymers and synthetic polymers demonstrate drastically different properties and hence completely different applications (Figure 1.1). Biopolymers such as DNA and proteins are constructed by biosynthetic pathways with strictly defined sequences, enabling an enormous number of distinct molecular species to be generated from a rather small set of building blocks. The vast sequence space that the biopolymers can access allows the living systems to adopt a unique solution to the considerable amount of challenges to sustain life and adapt to the changing environment: evolution. The complicated three-dimensional folded structures of many nucleic acids and proteins, and their unparalleled activities in catalyzing and regulating biochemical processes, are almost certainly a consequence of their ability to evolve in the form of sequence-defined polymers. In contrast, although much more structural and functional flexibility are allowed for synthetic monomers, the distinct molecular structures of the polymers made of these synthetic monomers remain modest compare to the highly diverse biopolymers. The discrepancy, having limited the use of synthetic polymers to mainly as bulk materials, derives from the lack of sequence control during chemical polymer synthesis. Unlike the biosynthetic machinery such as DNA polymerase or ribosome, which precisely controls

the addition of single monomer units by means of highly regulated enzymatic systems, the traditional chemical polymerization strategies, or even the contemporary developments of these strategies such as controlled radical polymerization, rely on relatively simple chemical transformations that are less specific and error-prone (compared to the high efficiency and high fidelity of the biopolymer synthesis). Although for a limited number of close synthetic analogs of biopolymers, solid-phase methods based on stepwise monomer incorporation have enabled sequence-defined synthesis, polymers that originate from solid-phase synthesis have not been compatible with directed evolution strategies to yield molecules with complex structural or functional properties. For these reasons, the control of single monomer addition using chemical polymerization means still remains largely aspirational.²

Biopolymer vs. Synthetic Polymer

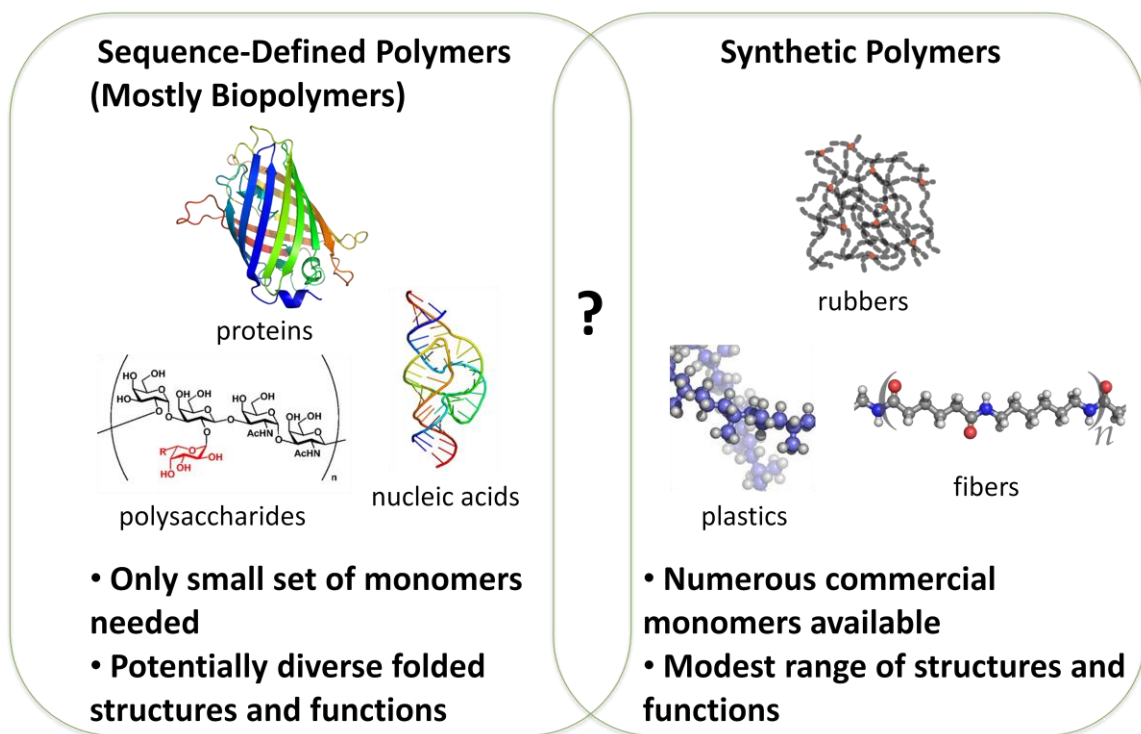


Figure 1.1. Although both belong to the category of polymers by definition, current biopolymers and synthetic polymers are drastically different macromolecules. Introducing defined sequence to synthetic polymers may open unprecedented structural and functional space for these man-made macromolecules.

1.2 Sequence-Defined Synthetic Polymers: The Next Holy Grail of Polymer

Chemistry.

Despite of the major structural and functional differences between biopolymers and synthetic polymers, the approaches to chemically generate these polymers share many common features. Biopolymers such as DNA and polypeptides can be readily synthesized by chemical means, i.e. solid phase synthesis. It has been well-established

that these synthetic biopolymers function no differently from those generated by biosynthetic processes.³ On the other hand, synthetic polymers can also be synthesized in a sequence-defined fashion by solid phase synthesis.⁴⁻⁷ Importantly, some of these sequence-defined synthetic polymers, such as β -peptides and various other peptidomimetics, have demonstrated folding properties that had previously been thought unique for biopolymers.^{7,8} Some particular sequences of these polymers are known to adopt secondary, tertiary, and even quaternary structures.⁹⁻¹² These folded synthetic polymers, just like folded biopolymers, may possess the potential to exert bioactive functions, such as binding to specific molecular targets or catalyzing reactions.¹³⁻¹⁸ Therefore, realizing precise sequence control in polymer chemistry may potentially enable unprecedented structures and functions of synthetic polymers, opening up application prospects that lie beyond the scope of current synthetic polymer materials.

Current developments of chemical polymerization strategies have already sought to achieve sequence control. Taking advantage of the tendency of cross-propagation of an electron donor and acceptor pair during a living radical polymerization, Lutz and coworkers achieved accurate positioning of maleimide functional groups along a polystyrene main chain.¹⁹ In this example, the electron-rich styrene is in large excess and its polymerization is the dominant reaction. During the course of styrene polymerization, small amounts of electron-deficient *N*-functionalized maleimide are added. Due to the high cross-propagation rate of this electron donor-acceptor pair in radical polymerization, maleimide monomers are incorporated in an alternating fashion within a narrow range of the growing polymer chain. By varying the addition time points,

maleimide with different functional side chains can be incorporated accurately at various positions of the polymer to achieve a sequence of the functional side chains (Figure 1.2a). A second sequence control strategy is represented by the work of Hillmyer and coworkers, in which a predefined tetrasubstituted cyclooctene monomers are polymerized in a regioselective ring-opening olefin metathesis polymerization (ROMP) reaction.²⁰ Since the sequence of all functional groups are already predefined on the monomer, regioselective ROMP would enable the concatemerization of these functional side chains to form sequence-regulated polymers (Figure 1.2b). Despite of these progresses, major hurdles still exist to achieve precise single monomer addition during the course of chemical polymerization. The product of these chemical polymerization reactions are still statistical polymers, lacking precise positioning of the functional group and a uniform molecular weight.

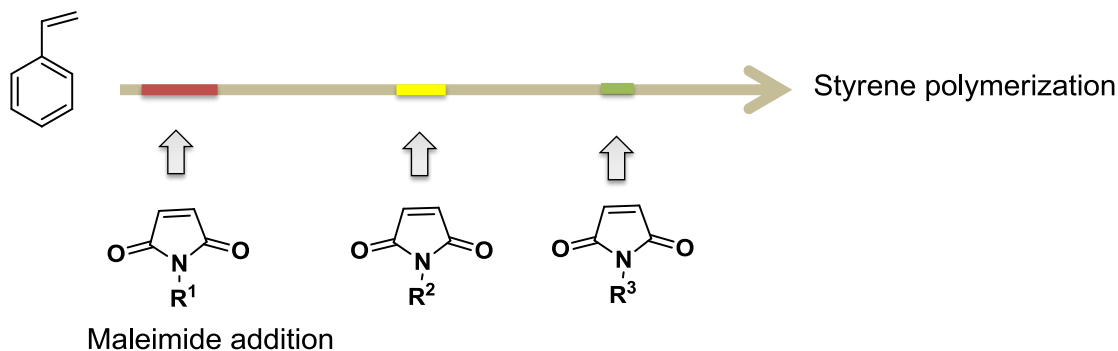
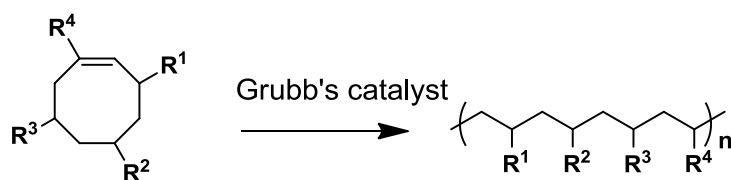
a**b**

Figure 1.2. Representative sequence control approaches in chemical polymerizations. (a) Small amount addition of modified maleimide monomers during a styrene polymerization allows fast cross-propagation of these two monomers and thus narrow distribution of maleimide functional groups on the polystyrene main chain. (b) Sequence-regulated polymer made by ROMP using cyclic alkene monomers with predefined side-chain pattern.

The best and probably the only route of polymer chemical synthesis to achieve absolute control over monomer sequence is the solid phase synthesis.²¹ In such a process, starting from an insoluble solid support, functional monomers bearing variable side chains can be sequentially connected one by one in a stepwise fashion. The unreacted coupling reagents or catalysts are washed away, and the deprotection step following chemical coupling reveals a new reactive end for another round of synthesis cycle. Since

its advent, the solid phase methodology has been primarily used to make natural form of polypeptides using α -L-amino acids as monomers,²¹ or natural form of DNA or RNA using the corresponding nucleoside phosphoramidites,²² but the versatility of this chemistry means that it can be used to couple other types of synthetic monomer structures in principle. In their work to synthesize peptoid, Zuckermann and coworkers has exemplified this chemical flexibility of the solid phase methodology.^{4,23} They divide the full cycle addition of an *N*-substituted glycine monomer into two submonomer coupling steps. In the first step, a α -bromoacetic acid is coupled to the solid support; this is followed by the substitution of the bromide group with a primary amine with diverse functionalities. The submonomer protocol eliminates the need of main-chain protecting groups, making it well suited to generate long polymers as large as 100-mers. One drawback of the solid phase approach is that with the growth of polymer, the reactive sites for coupling incoming monomer units tend to have more restricted accessibility which leads to the lower coupling yields. A potential solution to this problem is a protein chemical total synthesis strategy first developed by Kent and coworkers.²⁴ In such an approach, fragments of a synthetic polymer will be first synthesized on the solid support and then cleaved as unprotected peptides or with minimal side-chain protections. Subsequently the fragments can be connected by various coupling chemistries such as native chemical ligation to form the full-length protein.²⁵

In the current chemical approaches to the generation of sequence-controlled or sequence-defined synthetic polymers, a critical question has been left largely unanswered is which polymer sequence, if such sequences ever exist, will confer three-dimensional

folded structures or bioactive functions such as binding to specific molecular targets or catalyzing chemical reactions. Unlike biopolymers, the field of sequence-defined synthetic polymers is rather underexplored. Lacking prior knowledge on the underlying relationship between synthetic polymer sequence and its structure and functions, rational design of such polymers would be forbiddingly difficult. To address this problem, some researchers turn to Nature for inspirations.

1.3 Laboratory Directed Evolution: Towards Synthetic Polymers With Tailor-Made Function

Nature's approach in finding the functional solution from the astronomically large sequence space relies on evolution. The ability to evolve in the molecular contexts in living systems is one of the most significant factors for dominance of nucleic acids and proteins. During an evolution, a biopolymer with favorable functions such as binding to a certain signaling factor in a transduction pathway, or to catalyze a biochemical transformation that is critical to the thrive of the host living system, will increase the probability of the host's survival under natural selection pressure, hence greatly increasing the chance of the genetic information that encodes the synthesis of this biopolymer to be inherited by the next generation (Figure 1.3a). In this process, the ability of biopolymers to be translated from an information carrier molecule, i.e. DNA or RNA, is critical, because these information carrier molecules are able to replicate to pass the genetic information they carry to next generations, and they are also able to mutate to allow the exploration of greater sequence space towards evolution guided directions.

Through evolution, the living systems have been able to find molecular solutions to intriguing challenges facing life, and quickly become adapted to solve new challenges when environment changes.

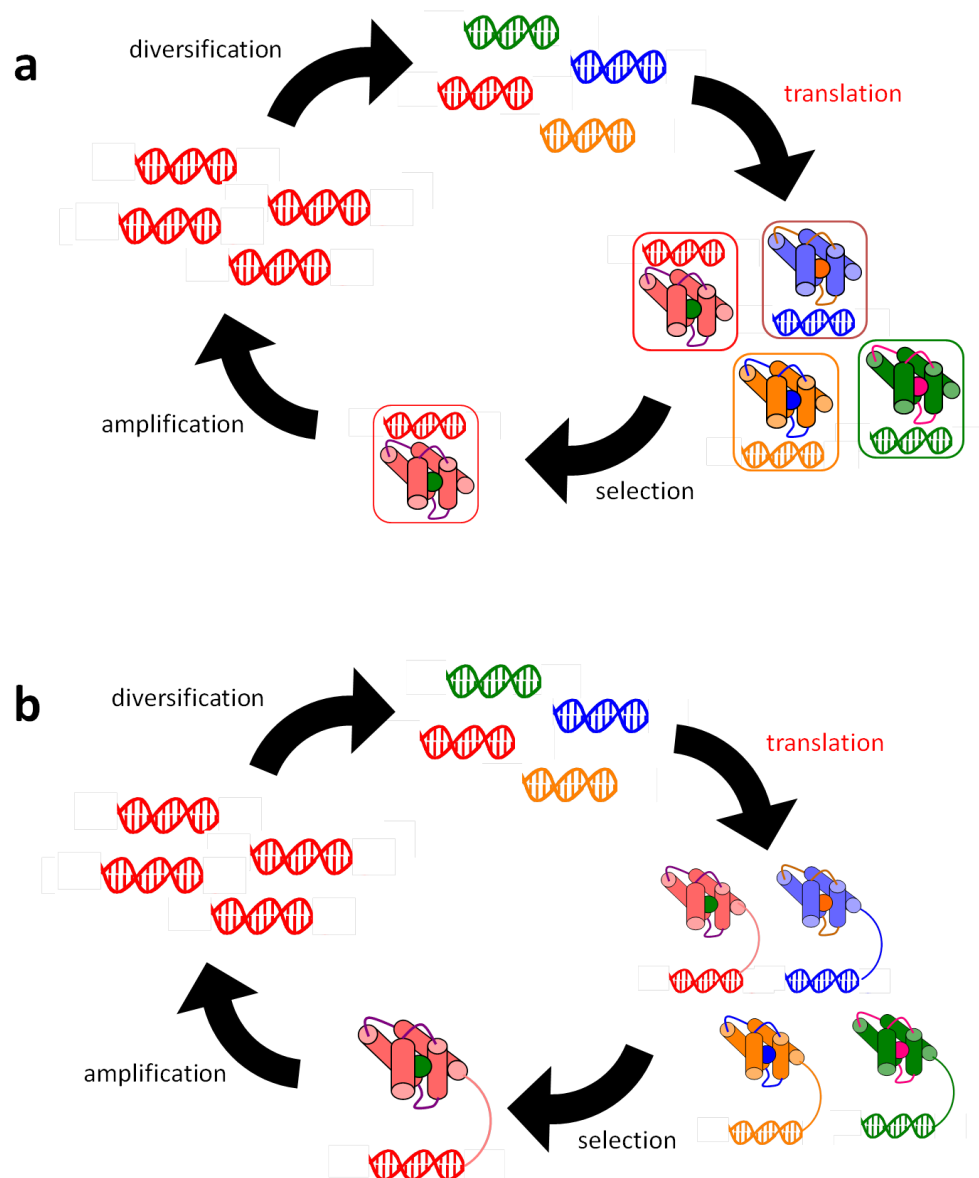


Figure 1.3

Figure 1.3 (Continued)

Figure 1.3. Schematic illustration of evolutions in Nature and in laboratory. (a) In Nature, biopolymers are translated from genetic information carrier molecules such as DNA. These biopolymers undergo selection based on their functional properties, and thus help their host to survive in the natural selection. The genetic information of these fit biopolymers is further amplified by the host's propagation, and sustains mutations to result the next generation of biopolymers for further selections. (b) Laboratory directed evolution (shown as *in vitro* evolution) inspired by the natural evolution process. A translation strategy that not only can generate polymer from its encoding gene, but also establishes the physical connection between the genotype and the phenotype is required to ensure successful directed evolution.

Inspired by Nature, laboratory directed evolutions have successfully generated artificial nucleic acids or proteins that possess tailor-made structures and functions (Figure 1.3b). Since the initial work on Systematic Evolution of Ligands by Exponential Enrichment (SELEX) on RNA,^{26,27} directed evolution has been used to enrich small numbers of functional biopolymers from a large library of random sequences. Directed evolution can be performed either in living systems (*in vivo*) or independent of a cellular environment (*in vitro*), with the latter also known as *in vitro* selection. The power of the *in vivo* directed evolution has been exemplified by numerous successful protein evolution works, in which proteins were engineered towards structures and functions beyond the scope of their natural properties. Recently, a continuous *in vivo* evolution system developed by Liu and coworkers has further highlighted the extensive evolvability of proteins towards a user-defined function.²⁸ In this system, the phage life cycle is used to complete all critical steps of a direction evolution: diversification, selection, and amplification, enabling iterative rounds of selection take place in an automated fashion. A

T7 RNA polymerase that recognizes a promoter sequence that is completely orthogonal to the wild type promoter has been successfully evolved using this system.²⁹

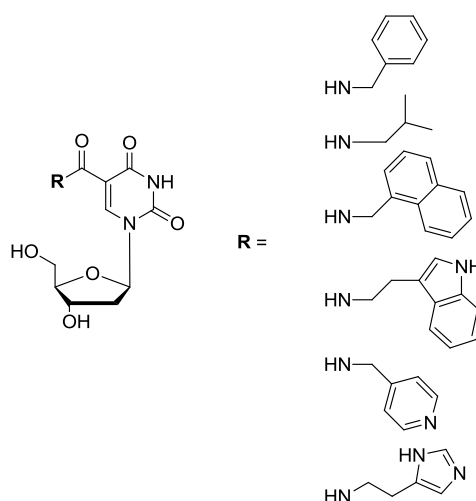
A major challenge in applying *in vivo* directed evolution system to evolve macromolecules with desired functions is the limitation of these systems to natural building blocks. Engineering a biosynthetic machinery that can readily use non-natural building blocks to construct functional polymers *in vivo* raises various additional challenges, ranging from the delivery of these non-natural building blocks into cell to the incorporation orthogonality of these building blocks within the cellular environment. Towards solving these problems, significant advancements have been achieved, represented by the *in vivo* suppression of amber codon of translation by unnatural amino acids with the mediation of evolved tRNA/aminoacyl tRNA synthetase pair. Demonstrated in the seminal works of Schultz,³⁰ and Tirrell,³¹ the engineered cellular translational machinery can incorporate a plethora of amino acids with unnatural side chains into proteins, such as *p*-azidophenylalanine, homopropargylglycine, and sulfotyrosine, thereby laying the foundation of directed evolution of these proteins containing unnatural amino acids. Limitations in generating functional proteins bearing unnatural amino acids include difficulty in engineering the translational machinery, reduced translation efficiency, restriction to incorporating only small subset of functional side chains, and only incorporating one type of unnatural amino acid in one protein sequence.³² Much work still need to be done before greater freedom can be achieved to evolve *in vivo* functional polymers bearing unnatural structural moieties.

In contrast, because biosynthetic machineries functioning in the cell-free environments pose much less restrictions on the type of building blocks, efforts on *in vitro* selections on functional synthetic systems translated from DNA or RNA are more fruitful. Both polymerase enzymes and ribosome have been used to generate biopolymers or their analogues with synthetic backbone or side chain structures, which are further evolved in *in vitro* selections for desired functions.

The polymerase enzyme-mediated polymerization of nucleotide triphosphates with altered base or sugar backbone structures has been subjected to considerable amount of research. Aptamers containing these nucleobase- or backbone-modified nucleotides have been evolved using *in vitro* selection approaches. Eaton and colleagues found several 5-position modified dUTP derivatives were able to be incorporated in a primer extension reaction mediated by Deep Vent DNA polymerase.³³ By generating a library of DNA with these 5-modified dUTPs incorporated in place of thymidines, they successfully evolved aptamers to bind strongly tumor necrosis factor receptor superfamily member 9 (TNFRSF9), a protein known to be difficult to raise DNA aptamers against (Figure 1.4a). Recently, Hirao and colleagues evolved aptamers with expanded genetic code: in addition to the four natural bases a fifth unnatural nucleobase, 7-(2-thienyl)imidazo[4,5-b]pyridine (Ds) was incorporated.³⁴ Ds base-pair with another unnatural base, a diol-modified 2-nitro-4 propynylpyrrole (Px) with high specificity, and this base pair is orthogonal to the four natural bases (Figure 1.4b). PCR with natural nucleotides supplemented with the extra Ds-Px nucleotide pair works almost as efficiently as normal PCR, resulting a nucleic acid polymer containing all natural bases

and Ds after strand separation. It was then determined that the aptamers with the extra Ds base evolved to bind VEGF-165 and IFN- γ have significantly stronger binding to the proteins compared to the conventional ones with only four natural bases. The unnatural Ds base plays an important role in binding, as mutating them back to adenine weakens the binding by up to a hundred fold.

a



b

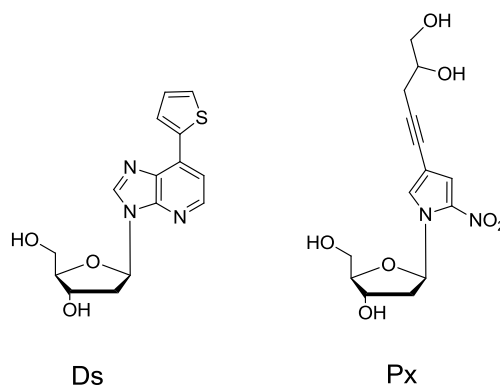


Figure 1.4. (a) Modified dUTP structures incorporated in the DNA aptamer evolved to bind TNFRSF9. (b) The structures of Ds and Px nucleotides that can form orthogonal base pair in PCR. The Ds base was incorporated in the evolved aptamers against VEGF-165 and IFN- γ , which showed a stronger affinity to the targets compared to the aptamers only containing natural bases.

First discovered by Chaput, Szostak, and coworkers^{35,36} and later significantly improved by the Chaput group,^{37,38} the Terminator DNA polymerase under optimized conditions has broad tolerance of base- and backbone-modified nucleotide substrates during a polymerization. For instance, it can polymerize a (3', 2')- α -L-threose nucleic acid (TNA) polymer in the presence of DNA template. An *in vitro* selection system was developed by the Chaput lab, using Terminator DNA polymerase to translate a library of DNA templates into TNA polymers (Figure 1.5a). Following the translation, a primer extension over the DNA template displace the TNA strand with a complementary DNA strand of the template, liberating and displaying the TNA polymer for *in vitro* selections. From this library a TNA aptamer was successfully selected to bind thrombin.³⁷ Perhaps the most impressive biosynthetic system that can generate synthetic genetic polymers with the ability of undergoing directed evolution is the XNA polymerase – engineered DNA polymerase that can synthesize polymers using synthetic nucleotide analogues (XNA) – evolved by Holliger and coworkers (Figure 1.5b).³⁹ Using a technology called Compartmentalized Self-Tagging (CST) which selects an encoding gene of a polymerase based on its ability to extend a primer using synthetic nucleotide analogues, they evolved enzymes from Terminator DNA polymerase to incorporate various XNAs including TNA, hexitol nucleic acids (HNA), 2'-O, 4'-methylene- β -D-ribonucleic acid (locked nucleic acids, LNA), cyclohexyl nucleic acid (CeNA), arabinonucleic acid (ANA), and 2'-fluoro-arabino-nucleic acid (FANA). Importantly, they also successfully evolved polymerase enzymes that can “reverse-transcribe” multiple XNAs back to DNA, thereby

completing the full cycle of information flow to enable iterative rounds of *in vitro* selection.

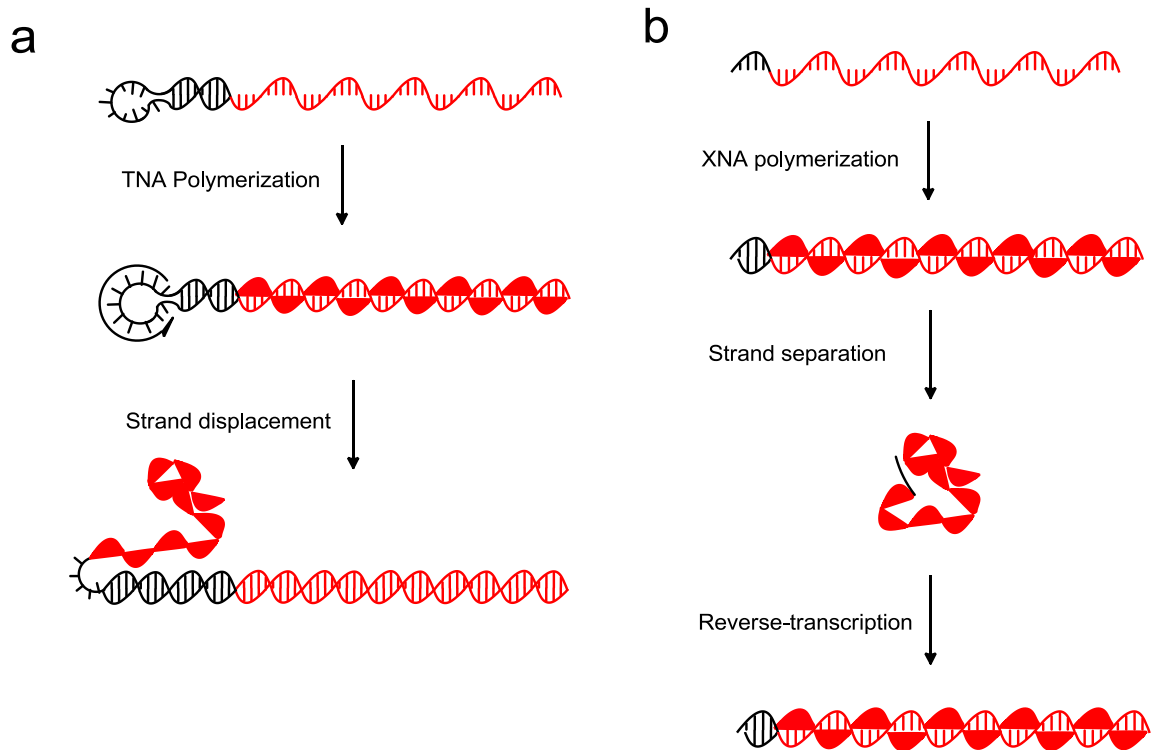


Figure 1.5. Two strategies to generate non-natural nucleic acid polymers for *in vitro* selection. (a) Translation of a DNA template into non-natural nucleic acid polymer is followed by a primer extension to displace the polymer product strand from the template strand. (b) Reverse-transcription is enabled by engineered polymerases. Therefore no template-product linkage is needed for the selection to be conducted.

Collectively, the works to use natural or engineered DNA polymerases to generate nucleic acid polymers containing synthetic base or sugar backbone modifications have represented some major breakthroughs in search of synthetic macromolecules with tailor-made functions; their development will further advance

peoples' understanding of the underlying relationship between polymer sequence and their structures and functions.

To expand the structural scope of synthetic macromolecules beyond nucleic acid polymers, researchers eyed on ribosome to produce polypeptides with synthetic building blocks.⁴⁰ Unlike the *in vivo* amber suppression system described above, the *in vitro* translation system, as known as Protein Synthesis Using Recombinant Elements (PURE), allows reassignment of sense codons to unnatural amino acids and can incorporate multiple unnatural amino acids in the same peptide.⁴¹ First developed by Ueda and coworkers, the PURE system has become the foundation of multiple polypeptide polymer *in vitro* selection systems, including Ribosome display,⁴² mRNA display,⁴³ and cDNA display.⁴⁴ Notably, the combination of the mRNA display system and “flexizyme”, a ribozyme developed by Suga and coworkers that can charge a tRNA with an arbitrary natural or unnatural amino acid, has opened a new research realm to synthesize and evolve polypeptide polymers with multiple unnatural amino acid building blocks.⁴⁵ Using this strategy, Szostak and colleagues recently evolved cyclic unnatural peptide aptamers against thrombin with the K_d of 4.5 nM.⁴⁶ The affinity is highly dependent on the unnatural side chains and the cyclized structure, with both the linear peptide and the one with all unnatural side chains mutated back to the natural counterparts have significantly reduced affinity to target.

Although many non-natural polymers have been generated by natural or engineered enzymatic systems and their subsequent directed evolutions raised aptamers

and catalysts, all examples to date still require close resemblance to natural biopolymers. Most enzymatic systems incorporate the unnatural building blocks with reduced efficiency, and remain unable to incorporate monomers with structures considerably different from the natural substrates. Efforts to expand the scope of synthetic monomers that can be accepted by the enzymatic systems have resulted significant progress,^{47–50} and will still remain active to bring more breakthroughs. In the same time, progress in this field also inspires research to design artificial systems that mimic, but not directly involve biosynthetic machineries to create functional synthetic polymers.

1.4 Enzyme-Free Translation of Nucleic Acids into Sequence-Defined Synthetic Polymers.

An alternative approach that parallels the use of biosynthetic pathways to generate sequence-defined synthetic polymers is to polymerize synthetic monomeric building blocks under the guidance of a DNA or RNA template. This process is also known as “translation”, since it translates the sequence information of nucleic acids into a corresponding polymer sequence. But different from more widely known ribosomal translation, the nucleic acid-templated chemical translation does not involve using ribosome or other biosynthetic machineries. The driving force of the translation fidelity is the specific Watson-Crick base-pairing interaction between the template and the reagent (or intermediate). Importantly, such translation systems do not rely on the monomeric building blocks’ ability to be accepted by an enzymatic process, thereby enabling the synthesis and potential laboratory directed evolution of sequence-defined synthetic

polymers that are structurally significantly different from the natural biopolymers. That said, these chemical translation systems usually require a different substantial set of criteria, many of which limit the scope of resulting polymer products as well as the translation efficiency. One common criterion is the ability of the monomeric building block, hence the product formed by the polymerization of these monomers, to hybridize sequence-specifically to the template. This requirement has limited the product of early chemical translation systems to mostly nucleic acid polymers.⁵¹

Orgel and coworkers were the first to study enzyme-free template-directed synthesis of oligoribonucleotides.^{52,53} They found in the presence of polynucleotide templates, the complementary mononucleotides tend to condensate to form oligomers when activated by carbodiimide, while the non-cognate nucleotides did not condensate under the same conditions.⁵² Although the initial goal of these studies was to understand how ribonucleic acid polymers originated in the prebiotic conditions, researchers eventually found that this templated polymerization is not an exclusive property of natural nucleic acids. In 1999, to understand the dominance of ribose and deoxyribose in Nature, Eschenmoser and coworkers systematically studied structural alternatives of ribose to form nucleic acid polymers and their properties in forming Watson-Crick base-pairing with natural nucleic acids.⁵⁴ This work has led the discovery of many non-natural nucleic acids that are able to form stable duplexes with DNA and RNA, inspiring others to study their abilities to polymerize in the presence of a complementary polynucleotide template. Orgel and coworkers reported templated polymerization of HNA and ANA (Figure 1.6a),⁵⁵ while Nielsen and colleagues pioneered the study of polymerization of

peptide nucleic acids (PNA) under the guidance of either natural nucleic acid polymer templates⁵⁶ or another PNA polymer (Figure 1.6a).⁵⁷ The studies on PNA arouse broad interests because unlike many other nucleic acids which have sugar backbone and negatively charged phosphodiester linkage, the backbone structure of PNA resembles more to that of protein and does not have charges. It is postulated that PNA may have played a role in the origin of nucleic acids as a functional biomacromolecule, prior to the advent of ribonucleic acids.⁵⁸ Furthermore, since PNA is achiral, using it to serve as a polymerization template raises interesting possibility of mirror-image nucleic acid products.

Despite the exciting results, these early templated polymerization reactions were barely efficient to generate short oligomers. Further developments were made possible by Lynn and coworkers, who reasoned that a reversible covalent linkage between nucleotide monomers may favor the formation of longer oligomers with matched sequence to the template, since longer sequence-matched products will be kinetically trapped to remain bound to the template (Figure 1.6b).⁵⁹ A subsequent reaction to transform the reversible linkage to an irreversible one can stabilize the product's structure to form a long polymer. With this perception, Lynn and coworkers incorporated an aldehyde and an amine functional groups on the ends of trinucleotides and showed that these trimers readily ligate in the presence of a matched template.^{59,60} The imine linkage formed in the ligation can be further reduced by cyanoborohydrides to form secondary amines and stabilize the product structure. Furthermore, the ligation reaction showed lowered product inhibition after reduction because the product with secondary amine linkages has reduced affinity to

the template compare to the intermediate with the imine linkages. Therefore, the templated ligation reaction can have multiple turnovers. The authors extended the ligation reaction to polymerizations of multiple short oligonucleotides and found they can achieve polymers as long as 32 nucleotides.⁶¹

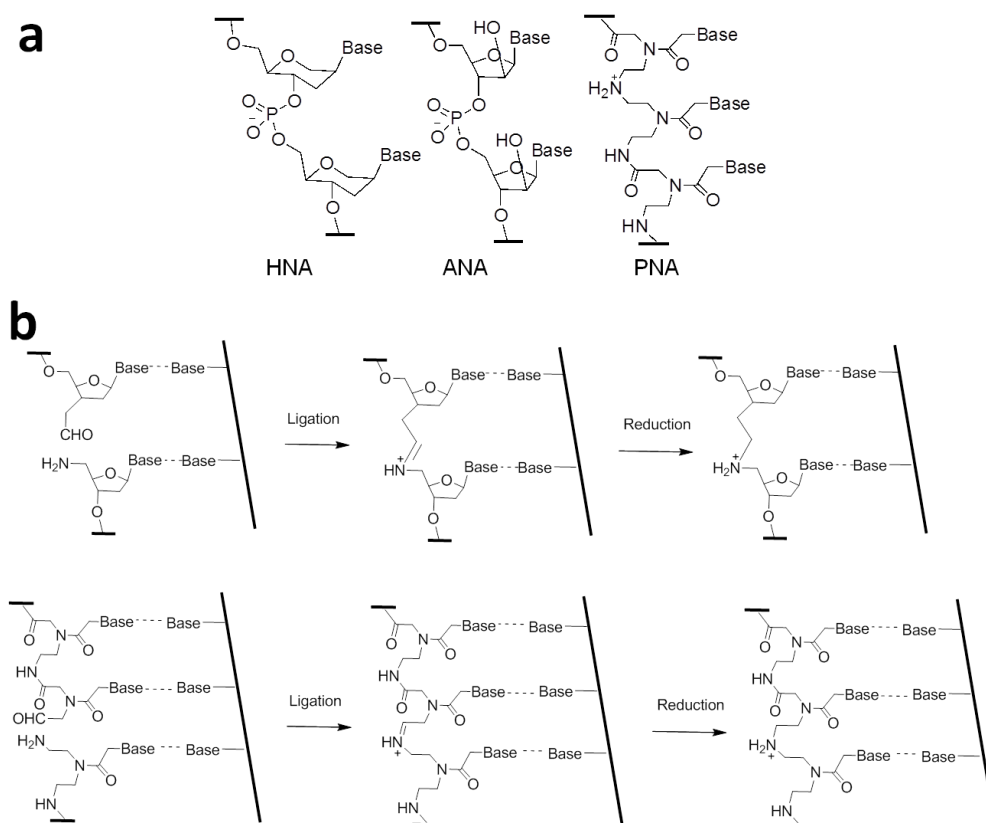


Figure 1.6. (a) Structure of nucleic acid polymers generated by enzyme-free nucleic acid-templated polymerization approach (b) Nucleic acid polymer templated ligation via reversible imine linkage, which is further reduced to form irreversible secondary amine.

Building on the works of Orgel, Lynn, and Nielsen, the Liu lab developed an enzyme-free DNA-templated translation strategy to translate DNA into PNA polymers.⁶²

Liu and coworkers have previously developed DNA-templated reactions in which bond-

forming reactions could readily occur between the functional groups on the respective template strand and reagent strand when these two strands hybridize.⁵¹ Inspired by Lynn's early work of DNA-templated polymerization of amido-DNA-aldehyde, former graduate students of the Liu lab Daniel Rosenbaum and Yevgeny Brudno took the lead to develop the translation system that can sequence-specifically polymerize short oligomeric PNA aldehyde in the presence of DNA templates.⁶² In such a system, PNA tetramer or pentamer with C-terminal aldehydes were first sequence-specifically anneal with the DNA template containing a 5'-amino group bearing hairpin structure. The N-terminal amino groups of the short PNA oligomers form imine linkages with the C-terminal aldehyde groups of the adjacent PNA oligomer, thereby achieving polymeric intermediates linked by imines. Subsequently reducing agents such as sodium cyanoborohydride are added to reduce the imines to secondary amines, stabilizing the linkages to irreversible covalent bonds (Figure 1.6b). The translation was proven to be highly sequence-specific, as a mismatched PNA oligomer cannot incorporate into the growing chain of PNA polymer and results a truncated product at the mismatched codon position. Brudno further expanded this translation strategy to generate a DNA-templated PNA polymer library, and demonstrated a mock *in vitro* selection could be completed to enrich a biotinylated member from a library containing 10^8 random non-biotinylated members by streptavidin pull-downs (Figure 1.7).⁶³ The successful enrichment of the biotinylated species by 10^6 fold clearly suggests an *in vitro* selection strategy powered by an enzyme-free translation system could enable the directed evolution of sequence-defined synthetic polymers beyond those generated by biosynthetic machineries.

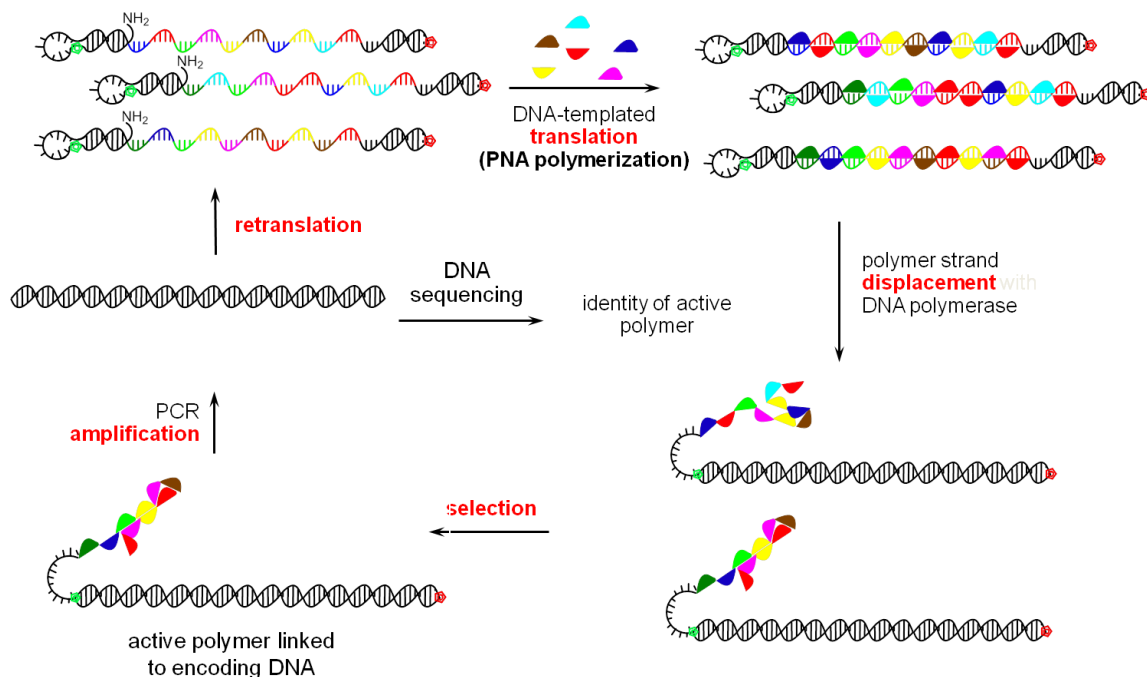


Figure 1.7. A full *in vitro* selection cycle of DNA-templated translation, strand displacement, selection, template re-generation, and retranslation to select for bioactive PNA polymers.

Although the recent advances in enzyme-free translation systems have established a strong foundation for the directed evolution of non-natural nucleic acid polymers, to date most of the systems still require the product to closely resemble the natural nucleic acids and maintain the ability to hybridize directly with a nucleic acid template. Towards an enzyme-free translation system with a more broad substrate scope, some meaningful attempts merit highlighting. O'Reilly, Turberfield, and coworkers,⁶⁴ and He and Liu⁶⁵ independently reported a strand displacement strategy that could facilitate multistep DNA-templated synthesis of oligomers that are structurally unrelated to nucleic acids. A key mechanism that allowed the successful synthesis of these

oligomers is after a monomeric building block on the reagent strand is transferred to the DNA template strand, a new incoming reagent strand displaces the old reagent strand by stronger hybridization with the template strand. Iterative strand displacements then allow the growth of the oligomeric product, with one monomer being added in a strand displacement cycle. However, as each new DNA reagent only adds one monomer, the formation of a full-length product requires frequent human intervention to add multiple reagents. Moreover, the differential affinities between the old reagent strand and the incoming reagent strand become less as the oligomer product grows longer, rendering the late stage coupling reactions less efficiently. This drawback significantly limits the potential length of the product to be made using the strand displacement strategies.

1.5 Thesis overview

To overcome the limitation of the current enzyme-free translation systems to only generating nucleic acid polymers, we sought to design and implement two translation systems that can significantly expand the structural and functional scope of evolvable synthetic polymers. The first system would enable enzyme-free translation of DNA into sequence-defined synthetic polymers structurally non-related to nucleic acids. Analogue to ribosomal translation, such a system would generate the sequence-defined polymers in a fully autonomous manner; no intervention from the researchers is needed to prepare the intermediates. This strategy should also enable translation of a DNA template library into a library of sequence-defined polymers with high fidelity. The second system relies on DNA ligase, rather than polymerase enzymes, to mediate the

ligation of short oligonucleotide building blocks with dense non-natural modifications. The density and chemical scope of the side-chain modifications on the nucleic acid polymers translated from DNA would far exceed the current systems, with even more potential for further expansion. We would seek to incorporate these two translation systems in a full cycle of translation, full-length product purification, selection, template re-generation, and re-translation, thereby demonstrating its capability to support *in vitro* selection of sequence-defined synthetic polymers for bioactive functions such as binding and catalysis.

Chapter two describes the design and implementation of the translation system. We detail the design and synthesis of a PNA macrocyclic substrate, the key component of the enzyme-free translation system. We also discuss the screening of the coupling chemistries to polymerize the substrates in the DNA-templated fashion. We also describe a terminator strategy, in combination with polyacrylamide gel electrophoresis (PAGE), to help determine the sequence-specificity of the translation as well as the identification of the full-length product. Finally, we demonstrate that such a translation system can support the generation of sequence-defined synthetic polymers with different backbone and side chain structures, as well as polymers with extended lengths.

Chapter three describes our efforts in designing a PNA genetic code that allows efficient and unbiased translation. The design and synthesis of the PNA macrocyclic substrates that incorporate diversified β -peptide building blocks are also described. In addition, we detail the re-optimization of the reaction condition of translation to facilitate

the efficient generation of a library of full-length β -peptide polymers. To prepare the β -peptide polymer library for *in vitro* selection, primer extension to introduce a complementary DNA strand to the template and full-length polymer purifications are undertaken to ensure the integrity of the selection results. Finally, we describe a mock selection experiment to enrich biotinylated species from a library of sequence-defined β -peptide polymers, demonstrating the integrated system is well-suited for future *in vitro* selections of functional synthetic polymers.

Chapter Four re-visits the strategy of using an enzymatic system to create sequence-defined nucleic acid polymers. But unlike the previous works that use polymerase enzymes, our strategy takes advantage of the versatility and broad substrate scope of T4 DNA ligase and use this enzyme to mediate the ligation of short oligonucleotides with various different non-natural side-chain modifications. We described the incorporation of eight different functional groups encoded by eight trinucleotide codons. The translation system exhibits high sequence specificity and efficiency and could generate a polymer of 50 consecutive substrates. Finally, we demonstrate an *in vitro* selection system of highly functionalized nucleic acid polymers based on the DNA-mediated translation strategy.

References

1. Jenkins, A. D., Kratochvil, P., Stepto, R. F. T. & W., S. U. INTERNATIONAL , UNION OF PURE GLOSSARY OF BASIC TERMS IN POLYMER. *Pure Appl. Chem.* **68**, 2287–2311 (1996).
2. Badi, N. & Lutz, J.-F. Sequence control in polymer synthesis. *Chem. Soc. Rev.* **38**, 3383–90 (2009).
3. Gibson, D. G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52–6 (2010).
4. Zuckermann, R. N., Kerr, J. M., Kent, S. B. H. & Moost, W. H. Efficient Method for the Preparation of Peptoids [Oligo(N-substituted glycines)] by Submonomer Solid-Phase Synthesis. *J. Am. Chem. Soc.* **114**, 10646–10647 (1992).
5. Culf, A. S. & Ouellette, R. J. Solid-phase synthesis of N-substituted glycine oligomers (alpha-peptoids) and derivatives. *Molecules* **15**, 5282–335 (2010).
6. Hartmann, L. & Börner, H. G. Precision polymers: monodisperse, monomer-sequence-defined segments to target future demands of polymers in medicine. *Adv. Mater.* **21**, 3425–31 (2009).
7. Seebach, D., Gardiner, J. & Zu, C.-. β -Peptidic Peptidomimetics. *Acc. Chem. Res.* **41**, 1366–1375 (2008).
8. Patch, J. a & Barron, A. E. Mimicry of bioactive peptides via non-natural, sequence-specific peptidomimetic oligomers. *Curr. Opin. Chem. Biol.* **6**, 872–877 (2002).
9. Martinek, T. a. & Fulop, F. Side-chain control of beta-peptide secondary structures. Design principles. *Eur. J. Biochem.* **270**, 3657–3666 (2003).
10. Daniels, D. S., Petersson, E. J., Qiu, J. X. & Schepartz, A. High-resolution structure of a beta-peptide bundle. *J. Am. Chem. Soc.* **129**, 1532–3 (2007).
11. Daura, X. *et al.* The beta-peptide hairpin in solution: conformational study of a beta-hexapeptide in methanol by NMR spectroscopy and MD simulation. *J. Am. Chem. Soc.* **123**, 2393–404 (2001).
12. Qiu, J. X., Petersson, E. J., Matthews, E. E. & Schepartz, A. Toward beta-amino acid proteins: a cooperatively folded beta-peptide quaternary structure. *J. Am. Chem. Soc.* **128**, 11338–9 (2006).

13. Stephens, O. M. *et al.* Inhibiting HIV fusion with a beta-peptide foldamer. *J. Am. Chem. Soc.* **127**, 13126–7 (2005).
14. Horne, W. S., Boersma, M. D., Windsor, M. a & Gellman, S. H. Sequence-based design of alpha/beta-peptide foldamers that mimic BH3 domains. *Angew. Chem. Int. Ed. Engl.* **47**, 2853–6 (2008).
15. Raguse, T. L., Porter, E. a, Weisblum, B. & Gellman, S. H. Structure-activity studies of 14-helical antimicrobial beta-peptides: probing the relationship between conformational stability and antimicrobial potency. *J. Am. Chem. Soc.* **124**, 12774–85 (2002).
16. Sadowsky, J. D. *et al.* (alpha/beta+alpha)-peptide antagonists of BH3 domain/Bcl-x(L) recognition: toward general strategies for foldamer-based inhibition of protein-protein interactions. *J. Am. Chem. Soc.* **129**, 139–54 (2007).
17. Koyack, M. J. & Cheng, R. P. in *Methods Mol. Biol.* (Guerois, R. & Paz, M. L. de la) **340**, 95–109 (Humana Press Inc., Totowa, NJ, 2006).
18. Bautista, A. D., Appelbaum, J. S., Craig, C. J., Michel, J. & Schepartz, A. Bridged beta(3)-peptide inhibitors of p53-hDM2 complexation: correlation between affinity and cell permeability. *J. Am. Chem. Soc.* **132**, 2904–6 (2010).
19. Pfeifer, S. & Lutz, J.-F. A facile procedure for controlling monomer sequence distribution in radical chain polymerizations. *J. Am. Chem. Soc.* **129**, 9542–3 (2007).
20. Zhang, J., Matta, M. E. & Hillmyer, M. A. Synthesis of Sequence-Specific Vinyl Copolymers by Regioselective ROMP of Multiply Substituted Cyclooctenes. *ACS Macro Lett.* **1**, 1383–7 (2012).
21. Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of. *J. Am. Chem. Soc.* **85**, 2149–2154 (1963).
22. Reese, C. B. Oligo- and poly-nucleotides: 50 years of chemical synthesis. *Org. Biomol. Chem.* **3**, 3851–68 (2005).
23. Burkoth, T. S., Fafarman, A. T., Charych, D. H., Connolly, M. D. & Zuckermann, R. N. Incorporation of unprotected heterocyclic side chains into peptoid oligomers via solid-phase submonomer synthesis. *J. Am. Chem. Soc.* **125**, 8841–5 (2003).
24. Kent, S. B. Chemical synthesis of peptides and proteins. *Annu. Rev. Biochem.* **57**, 957–89 (1988).

25. Kent, S. B. H. Total chemical synthesis of proteins. *Chem. Soc. Rev.* **38**, 338–51 (2009).
26. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–22 (1990).
27. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–10 (1990).
28. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011).
29. Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. a & Liu, D. R. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nat. Chem. Bio.* **10**, 216–222 (2014).
30. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. & Schultz, P. G. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244**, 182–8 (1989).
31. Link, J., Mock, M. L. & Tirrell, D. Non-canonical amino acids in protein engineering. *Curr. Opin. Biotechnol.* **14**, 603–609 (2003).
32. Brustad, E. M. & Arnold, F. H. Optimizing non-natural protein function with directed evolution. *Curr. Opin. Chem. Biol.* **15**, 201–10 (2011).
33. Vaught, J. D. *et al.* Expanding the chemistry of DNA for in vitro selection. *J. Am. Chem. Soc.* **132**, 4141–51 (2010).
34. Kimoto, M., Yamashige, R., Matsunaga, K., Yokoyama, S. & Hirao, I. Generation of high-affinity DNA aptamers using an expanded genetic alphabet. *Nat. Biotechnol.* **31**, 453–7 (2013).
35. Chaput, J. C. & Szostak, J. W. TNA synthesis by DNA polymerases. *J. Am. Chem. Soc.* **125**, 9274–5 (2003).
36. Ichida, J. K. *et al.* An in vitro selection system for TNA. *J. Am. Chem. Soc.* **127**, 2802–3 (2005).
37. Yu, H., Zhang, S. & Chaput, J. C. Darwinian evolution of an alternative genetic system provides support for TNA as an RNA progenitor. *Nat. Chem.* **4**, 183–7 (2012).

38. Yu, H., Zhang, S., Dunn, M. R. & Chaput, J. C. An efficient and faithful in vitro replication system for threose nucleic acid. *J. Am. Chem. Soc.* **135**, 3583–91 (2013).
39. Pinheiro, V. B. *et al.* Synthetic Genetic Polymers Capable of Heredity and Evolution. *Science* **336**, 341–344 (2012).
40. Lipovsek, D. & Plückthun, A. In-vitro protein evolution by ribosome display and mRNA display. *J. Immunol. Methods* **290**, 51–67 (2004).
41. Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19**, 751–5 (2001).
42. Mattheakis, L. C., Bhatt, R. R. & Dower, W. J. An in vitro polysome display system for identifying ligands from very large peptide libraries. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 9022–6 (1994).
43. Roberts, R. W. & Szostak, J. W. RNA-peptide fusions for the in vitro selection of peptides. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12297–12302 (1997).
44. Yamaguchi, J. *et al.* cDNA display: a novel screening method for functional disulfide-rich peptides by solid-phase synthesis and stabilization of mRNA-protein fusions. *Nucl. Acids Res.* **37**, e108 (2009).
45. Ohuchi, M., Murakami, H. & Suga, H. The flexizyme system: a highly flexible tRNA aminoacylation tool for the translation apparatus. *Curr. Opin. Chem. Biol.* **11**, 537–42 (2007).
46. Schlippe, Y. V. G., Hartman, M. C. T., Josephson, K. & Szostak, J. W. In vitro selection of highly modified cyclic peptides that act as tight binding inhibitors. *J. Am. Chem. Soc.* **134**, 10469–77 (2012).
47. Loakes, D., Gallego, J., Pinheiro, V. B., Kool, E. T. & Holliger, P. Evolving a polymerase for hydrophobic base analogues. *J. Am. Chem. Soc.* **131**, 14827–37 (2009).
48. Pinheiro, V. B. & Holliger, P. The XNA world: progress towards replication and evolution of synthetic genetic polymers. *Curr. Opin. Chem. Biol.* **16**, 245–52 (2012).
49. Kawakami, T., Ishizawa, T. & Murakami, H. Extensive reprogramming of the genetic code for genetically encoded synthesis of highly N-alkylated polycyclic peptidomimetics. *J. Am. Chem. Soc.* **135**, 12297–304 (2013).

50. Josephson, K., Hartman, M. C. T. & Szostak, J. W. Ribosomal synthesis of unnatural peptides. *J. Am. Chem. Soc.* **127**, 11727–35 (2005).
51. Li, X. & Liu, D. R. DNA-templated organic synthesis: nature's strategy for controlling chemical reactivity applied to synthetic molecules. *Angew. Chem. Int. Ed.* **43**, 4848–70 (2004).
52. Sulston, J., Lohnmann, R., Orgel, L. E. & Miles, H. T. Nonenzymatic Synthesis of Oligoadenylates on a Polyuridylic acid template. *Proc. Natl. Acad. Sci. U.S.A.* **59**, 726–733 (1967).
53. Sulston, J. *et al.* Non-enzymatic Oligonucleotide Synthesis on a Polycytidylate Template. *J. Mol. Biol.* **40**, 227–234 (1969).
54. Eschenmoser, A. Chemical Etiology of Nucleic Acid Structure. *Science* **284**, 2118–2124 (1999).
55. Kozlov, I. A. *et al.* Nonenzymatic Template-Directed Reactions on Altritol Oligomers. *Chem. Eur. J.* **6**, 151–155 (2000).
56. Schmidt, J. G., Christensen, L., Nielsen, P. E. & Orgel, L. E. Information transfer from DNA to peptide nucleic acids by template-directed syntheses. *Nucleic Acids Res.* **25**, 4792–6 (1997).
57. Boehler, C., Nielsen, P. E., Orgel, L. E. & Böhrer, C. Template switching between PNA and RNA oligonucleotides. *Nature* **376**, 578–581 (1995).
58. Nelson, K. E., Levy, M. & Miller, S. L. Peptide nucleic acids rather than RNA may have been the first genetic molecule. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3868–71 (2000).
59. Goodwin, J. T. & Lynn, D. G. Template-directed synthesis: use of a reversible reaction. *Chem. Commun.* 9197–8 (1992).
60. Leitzel, J. C. & Lynn, D. G. Template-directed ligation: from DNA towards different versatile templates. *Chem. Rec.* **1**, 53–62 (2001).
61. Li, X., Zhan, Z.-Y. J., Knipe, R. & Lynn, D. G. DNA-catalyzed polymerization. *J. Am. Chem. Soc.* **124**, 746–7 (2002).
62. Rosenbaum, D. M. & Liu, D. R. Efficient and sequence-specific DNA-templated polymerization of peptide nucleic acid aldehydes. *J. Am. Chem. Soc.* **125**, 13924–5 (2003).

- 63. Brudno, Y., Birnbaum, M. E., Kleiner, R. E. & Liu, D. R. An in vitro translation, selection and amplification system for peptide nucleic acids. *Nat. Chem. Biol.* **6**, 148–55 (2010).
- 64. McKee, M. L. *et al.* Multistep DNA-Templated Reactions for the Synthesis of Functional Sequence Controlled Oligomers. *Angew. Chem. Int. Ed.* **49**, 7948–51 (2010).
- 65. He, Y. & Liu, D. R. A sequential strand-displacement strategy enables efficient six-step DNA-templated synthesis. *J. Am. Chem. Soc.* **133**, 9972–5 (2011).

Chapter Two

Enzyme-Free Translation of DNA into Sequence-Defined Non-Nucleic Acid Synthetic Polymers

Jia Niu, Ryan Hili, and David R. Liu

Ryan Hili helped to design the translation system and analyze the data.

Adapted from: Niu, J., Hili, R., and Liu, D. R. Enzyme-free translation of DNA into sequence-defined synthetic polymers structurally unrelated to nucleic acids. *Nature Chem.* 5, 282-292 (2013).

2.1 Introduction

An enzyme-free translation system to enable translation of DNA into sequence-defined synthetic polymers structurally unrelated to nucleic acids would offer unprecedented access to structural and functional spaces for synthetic polymers. Importantly, since such synthetic polymers are encoded by DNA, the information carrier molecule in all living systems, which has been subjected to substantial amount of studies, *in vitro* selection of these synthetic polymers and subsequent deduction of their sequence information would be able to seek technological assistance from well-established genetic techniques such as PCR and DNA sequencing.

We envision that to achieve the goal of translating DNA into a non-nucleic acid polymer suitable for iterative rounds of *in vitro* selection, the translation system should enable the following features: (1) adapter units that mediate the sequence-specific recognition of DNA template codons by the monomeric building blocks can be cleaved from the product with minimal perturbations on the structural integrity of the product; (2) compatible bioconjugation chemistry will allow efficient DNA-templated polymerization and the covalent linkage between the DNA template and the translation product; (3) translation products have structural elements sufficient to support polymer folding and function. Our previous studies on DNA-templated PNA polymerizations may offer important inspirations,¹⁻³ but the translation system proposed herein requires entirely new design of all major components and re-optimization of reaction conditions, as it differs significantly from all previous systems.

2.2 Translation Strategy Design and Substrate Syntheses.

Inspired by Nature, we sought to emulate the function of a transfer RNA (tRNA) as an adapter that recognizes a template codon and brings an amino acid monomer into reactive proximity of a growing peptide chain. Therefore our design of the core component of the translation system is based on a “substrate” – a molecular construct that interacts with the DNA template – comprising three parts: (1) mimicking a tRNA anticodon, a PNA pentamer to sequence-specifically recognize a DNA template codon; (2) mimicking the amino acid monomer of the charged tRNA, a synthetic polymer building block that bears no necessary structural relationship with nucleic acids; and (3) cleavable linkers that connect each PNA anticodon with its cognate synthetic polymer building block in a macrocycle (Figure 2.1b). The substrates are macrocycles to decrease the entropic penalty of the building block coupling reactions and to increase their regioselectivity by aligning otherwise freely rotating building blocks into conformations that facilitate reactions between functional groups.⁴

The translation process is designed to proceed in three stages (Figure 2.1a). First, substrates hybridize sequence-specifically to a DNA template that contains a 5' hairpin followed by consecutive DNA codons. Substrate-template hybridization increases the effective molarity of reactive groups on adjacent building blocks. Second, a catalyst or reagent initiates coupling between building blocks, resulting in their oligomerization in a sequence-programmed order. Since the 5' end of the DNA template contains a group capable of coupling with the first building block, the synthetic polymer emerges from the

translation process covalently linked to its encoding template. As a result, translation products can undergo iterative rounds of *in vitro* selection, template replication, and retranslation. Finally, after the oligomerization reaction is complete, the linkers between the PNA anticodons and the synthetic polymer are cleaved, releasing the linear synthetic polymer-DNA template conjugate from the PNA adapters. Because the entire translation process does not require any structural or functional feature of the synthetic polymer building blocks beyond their ability to support coupling and linker cleavage, this strategy should be compatible with a wide variety of polymers, including those unrelated to nucleic acids.

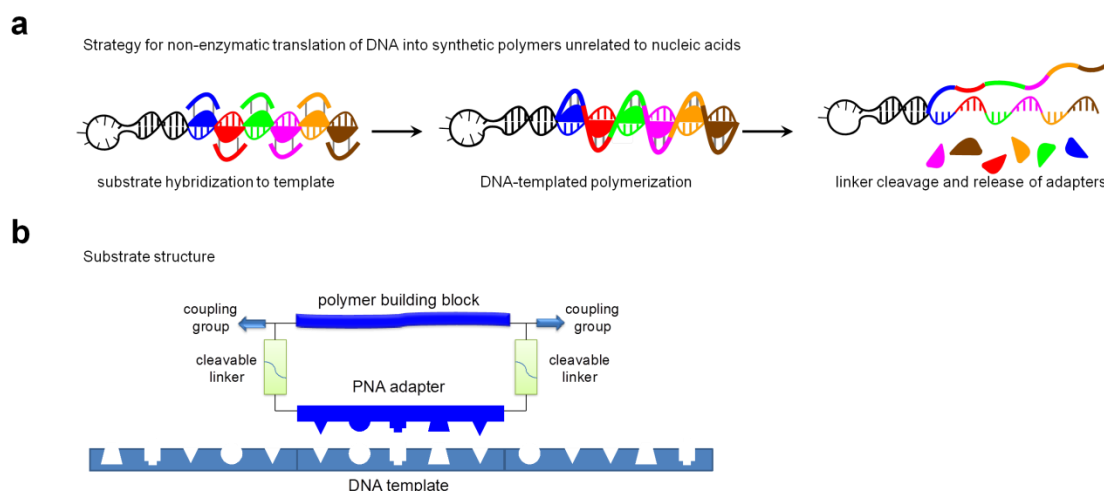


Figure 2.1. (a) Schematic illustration of the DNA-templated translation. Macrocyclic substrates hybridize with the codons on a DNA template, organizing synthetic polymer building blocks along the template. Coupling reactions then oligomerize these pre-organized substrates. Finally, linker cleavage releases the PNA adapters and liberates the synthetic polymer product. (b) Representation of a macrocyclic substrate.

We considered two substrate architectures: one design is featured with heterobifunctional ‘AB’ building blocks that each contains both types of coupling reaction functional groups, and the other design is featured with homobifunctional ‘AA’ or ‘BB’ building blocks that each contains only one type of reactive functional group (Fig. 2.2a). AB building blocks, when properly aligned, are capable of reacting with any adjacent AB building blocks, while AA or BB building blocks can only react with adjacent BB or AA building blocks, respectively. Although AB-type substrates are conceptually simpler and place fewer restrictions on the resulting polymer structures than AA/BB-type substrates, they may be prone to intramolecular cyclization, an undesired alternative to polymerization (Figure 2.2).^{5,6} We sought to test chemistries known to be compatible with DNA and with solid-phase peptide synthesis. The substrates with the corresponding functional groups were synthesized for the following five candidate coupling chemistries: amine acylation, reductive amination, oxime and hydrazone formation, and Cu(I)-catalyzed alkyne-azide 1,3-dipolar cycloaddition (CuAAC or ‘click’ chemistry).⁷

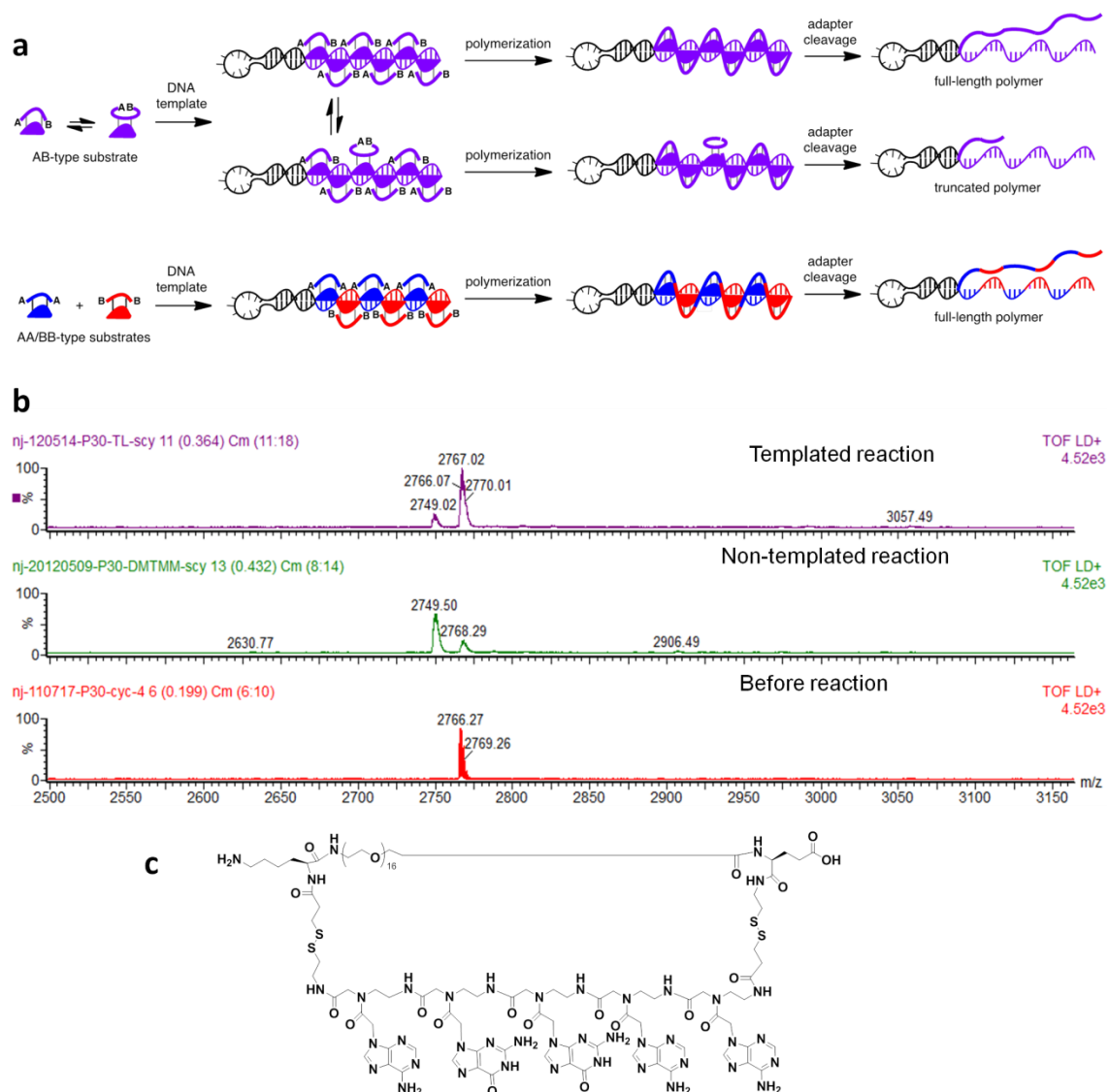


Figure 2.2. Intramolecular cyclization of the AB-type substrates. (a) Schematic illustration of AB-type and AA/BB type translation reactions. The AB-type substrates are more prone to intramolecular cyclization reaction, resulting in a truncated product. (b) Mass spectrometry analysis of an AB-type substrate, **20** (PNA sequence: AAGGA), after it was subjected to translation reaction conditions, both in the presence and in the absence of a 5-mer DNA template **T5** (5'-TCCTT-3'). Comparing the MALDI spectra before and after reaction, both the templated reaction and non-templated reaction showed a second peak of M-18, indicating that a cyclization reaction occurs by losing one molecule of water. The templated reaction has less intramolecularly cyclized product formed than the non-templated reaction, presumably due to the enhanced rigidity after hybridizing with the DNA template. (c) The structure of substrate **20**.

The syntheses of substrates were achieved by combining solid- and solution-phase reactions (Figure 2.3). With the exception of the substrates with C-terminal carboxylic acid functional group whose syntheses started from a Wang resin preloaded with Fmoc-Glu-ODmab monomer, the syntheses of the rest of the substrates all started from a Rink amide resin. Resins with low functional group loading were used because they tend to favor the cyclization and avoid multimerization during the on-resin cyclization reaction.⁸ For the syntheses using Rink amide resin, the first amino acid coupled introduces the functional group for coupling at the C-terminus, followed by a branch-point amino acid monomer Fmoc-Glu-ODmab to introduce a Dmab (4-{*N*-[1-(4,4-Dimethyl-2,6-dioxocyclohexylidene)-3-methylbutyl]-amino}benzyl, protecting carboxylic acid) orthogonally protected side chain for cyclization.⁸ The amino acid monomers for the synthetic polymer building blocks were subsequently coupled, followed by a monomer to introduce the N-terminal coupling group. A disulfide containing amino acid monomer, ((*N*-Fmoc-aminoethyl)-disulfanyl)-propanoic acid, was then coupled to introduce the C-terminal disulfide linker. Subsequently the PNA sequence was introduced by consecutive coupling of PNA monobases with the order from C-terminus to N-terminus. Lastly the N-terminal disulfide linker was coupled to finish the synthesis of the linear precursor. The entire synthesis of the linear precursor was completed on an automated peptide synthesizer. Subsequently the resin was placed in a glass reaction vessel. The branch-point Dmab-protected carboxylic acid side chain was deprotected by the treatment of hydrazine in DMF, and the revealed carboxylic acid was activated by diisopropylcarbodiimide (DIC) and *N*-hydroxybenzotriazole (HOBt) to

react with the N-terminal amine of the linear precursor on resin to form the cyclized product. After cyclization reaction and prior to cleavage of the product, some substrates require a selective deprotection of lysine side chain and a subsequent functional acid monomer coupling step to introduce the coupling functional group, enabled by the ability of ultra-acid-labile monomethyltrityl (MMT) protection of lysine side chains to be selectively deprotected by a weak acidic condition (1% TFA in DCM). The global deprotection and product cleavage was effected by a cleavage cocktail of 95% TFA, 2.5% water, and 2.5% triisopropylsilane (TIPS), with the later two ingredients as scavengers for the deprotected cations. The deprotected substrates were precipitated by diethylether and subsequently HPLC purified. Some substrates require an additional modification step after HPLC purification to conjugate a functional carboxylate NHS ester with a lysine side chain. For the initial studies to evaluate substrate architecture and coupling chemistries, polyethylene glycol (PEG) chains served as the synthetic polymer building blocks.

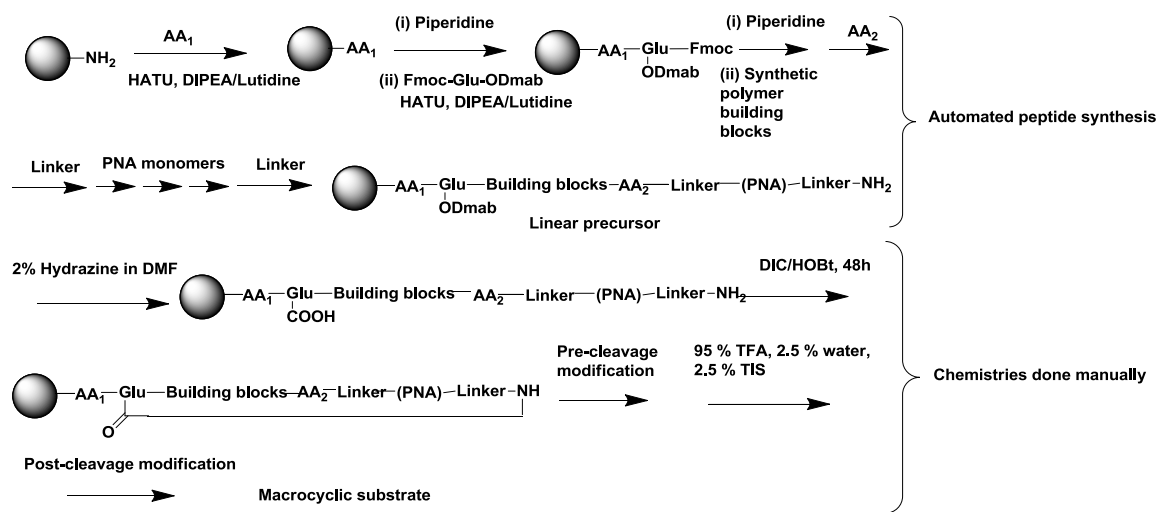


Figure 2.3. Synthetic procedure of substrates

2.3. Evaluation of Building Block Coupling Chemistries and Substrate Architectures

We set up translation experiments to evaluate substrate architectures and coupling chemistries by attempting the oligomerization of six consecutive building blocks on a DNA template containing six codons. For testing AB substrates, all six template codons were identical (TCCTT); for testing AA/BB substrates, templates contained alternating AA codons (AATCC) and BB codons (ATACC). These codons were chosen because they met the criteria of efficient and sequence-specific hybridization between DNA and PNA in our previous studies. Although G was not used in template codons in this particular experiment, in further studies we have verified that G containing codons are compatible with the translation strategy (see Chapter Three). After oligomerization of the substrates, dithiothreitol (DTT) was added to cleave the disulfide linkers and liberate the synthetic polymer products.

Translation reactions were analyzed by denaturing PAGE and generated up to seven distinct bands on the gel. We hypothesized that the fastest-migrating (lowest) band corresponds to the starting template, the highest band corresponds to the full-length translation product, and the intermediate bands are truncated intermediates (Figure 2.4a). When any of the three key components (the DNA template, the substrates or the coupling reagent or catalyst) were omitted from the reaction mixture, no products were observed (Figure 2.4b). Mass spectrometry (MS) analysis confirmed that the material from the

seventh (topmost) band has a molecular weight consistent with that of the full-length translation product (see below).

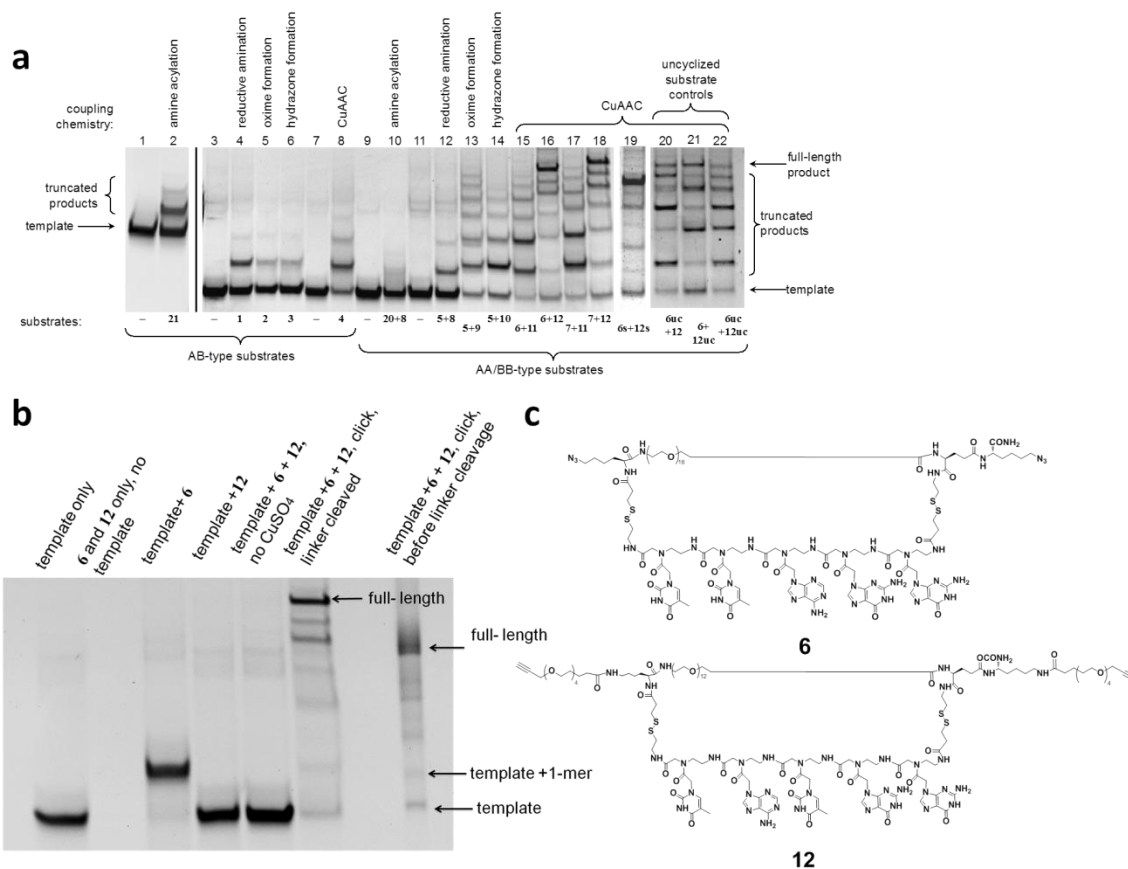


Figure 2.4. (a) Denaturing PAGE analysis of translation reactions to evaluate five candidate coupling chemistries. The templates used in lanes 1 and 2 contain eight codons; all other templates contain six codons. The denaturing 10% PAGE–urea gel was electrophoresed in TBE buffer at 200 V for 1 h at room temperature, then stained with ethidium bromide. AA/BB-type substrates coupled by CuAAC offered the highest overall yields of full-length translation products (lanes 16, 18 and 19). In contrast, AB-type substrates failed to yield full-length translation products (lanes 1–8), potentially due to competitive intramolecular cyclization. Notably, the macrocyclic nature of the substrate is required for efficient translation, as non-cyclized substrates showed significantly impaired oligomerization (lanes 20–22). The structures of the substrates other than **6** and **12** can be found in the Molecular Structures section. (b) DNA-templated translation using substrates **6** and **12**. Key ingredients of the translation reaction were intentionally omitted to confirm the formation of full-length translation product. (c) Structures of substrates **6** and **12**.

The DNA-templated oligomerization of AB- and AA/BB-type substrates using amine acylation as the coupling chemistry failed to couple more than two building blocks (Figure 2.4a, lanes 2 and 10). These results are consistent with previous findings that amine acylation is inefficient in DNA-templated oligomerizations.^{1,9,10} Moreover, the intramolecular cyclization of AB-type substrates significantly decreased product yields. Next, we tested reductive amination, which has been successfully used to polymerize analogues of DNA¹¹ and PNA.¹⁻³ Unfortunately, both AB- and AA/BB-type substrates coupled no more than twice under the conditions tested (Figure 2.4a, lanes 4 and 12). We hypothesized that the strict geometric requirements of reductive amination are problematic in the context of these large macrocyclic substrates.

We next explored chemistries that form products that are thermodynamically more stable than imines, such as hydrazone- and oxime-based coupling reactions. Although hydrazone and oxime formation using AB building blocks did not deliver higher levels of desired products due to efficient competing intramolecular cyclization (Figure 2.4a, lanes 5 and 6), oligomerization of AA/BB building blocks using these chemistries successfully coupled up to five building blocks on the template (Figure 2.4a, lanes 13 and 14). Despite these promising observations, we were concerned with the potential instability of a polymer containing oxime linkages in the basic pH conditions that may be needed during future *in vitro* selection procedures.¹²

Finally, we tested the ability of the CuAAC reaction to effect translation.^{13,14} Although inefficient coupling was again observed for AB-type substrates,

presumably due to competitive substrate cyclization (Figure 2.4a, lane 8), the CuAAC reaction resulted in the full-length hexamer product when using AA/BB-type substrates (Figure 2.4a, lane 15). Further optimization of building-block geometries significantly improved translation efficiency, so that up to 70% of translation products were full-length (Figure 2.4a, lanes 16, 18 and 19).

We determined the structural requirements for efficient translation using the CuAAC coupling chemistry. First, we varied the length of the synthetic polymer building block. DNA-templated oligomerization of substrates containing synthetic polymer building blocks as short as four PEG units was efficient in generating full-length product (Figure 2.5). Surprisingly, a moderate amount of full-length product is generated even when the synthetic polymer moiety is completely omitted (Figure 2.5). These results suggest that the DNA-templated translation strategy can accommodate a wide range of lengths of the synthetic polymer building blocks. We next varied the lengths of the groups between the CuAAC coupling partners and the synthetic polymer building blocks. Alkynyl substrates with spacers of one PEG unit and those with spacers of four PEG units both yielded full-length translation products efficiently (Figure 2.5), suggesting that the translation is relatively insensitive to the length of these spacers. Interestingly, the oxygen atom at the β -position of the alkynyl group is required for efficient translation (compare Figure 2.4a, lanes 15 and 17 with lane 19). We speculate that the increased polarity of the oxygen atom may help reduce the tendency of the otherwise hydrophobic spacers to adopt compact conformations that are incompatible with oligomerization.

Last but not least, we tested if the macrocyclic nature of the substrates improved translation outcomes. Uncyclized substrates **6-uc** and **12-uc**, the linear analogues of **6** and **12**, result in prematurely terminated oligomerization when co-polymerized with macrocyclic **12** and **6**, respectively (Figure 2.4a, lanes 20 and 21). When both **6-uc** and **12-uc** were used together in a DNA-templated translation reaction, the yield of full-length product was much lower than when macrocyclic substrates **6** and **12** were used (Figure 2.4a, lane 22 versus lane 16). In light of these findings, we pursued the CuAAC-mediated oligomerization of AA/BB-type macrocyclic substrates **6** and **12** (Figure 2.4c) as our model system for development.

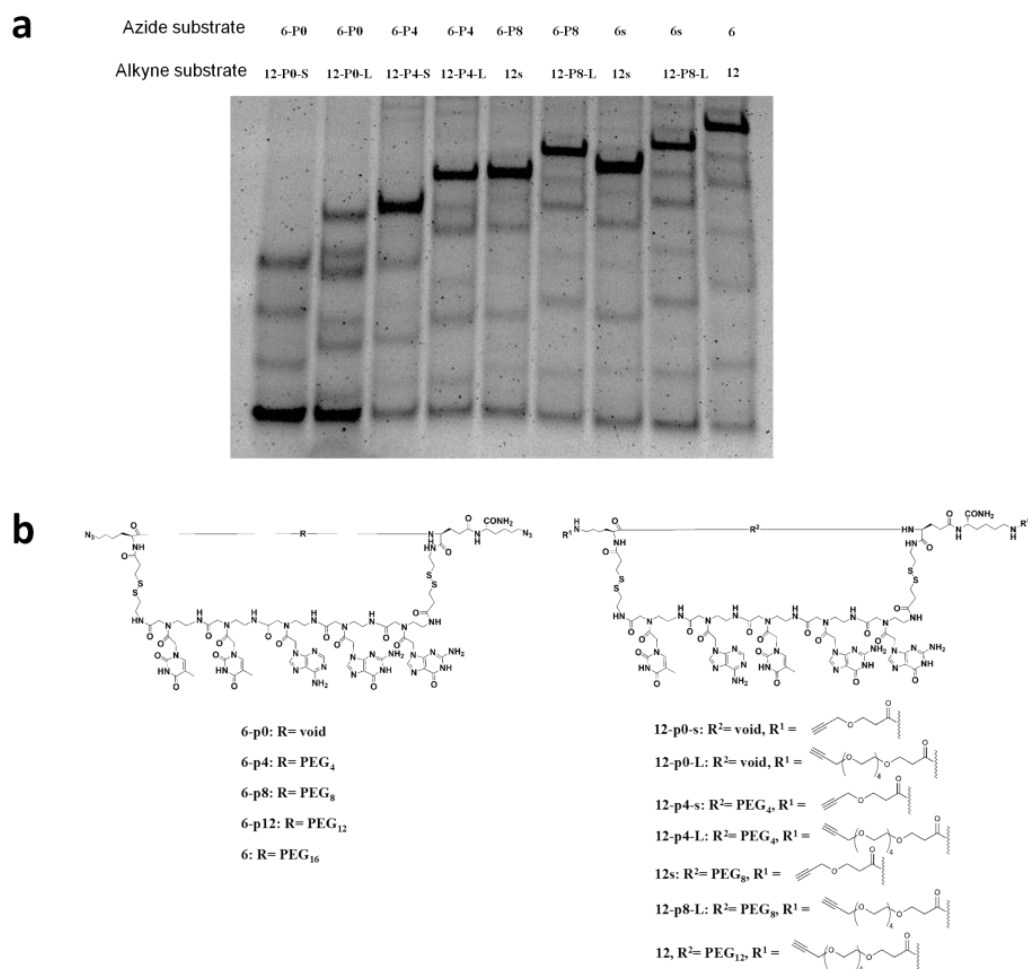


Figure 2.5. DNA-templated translation using substrates with different lengths of synthetic polymer building blocks and spacers. (a) Denaturing PAGE analysis of translation products. (b) The structures of substrates used in this experiment.

2.4 Characterization of the Translation Product.

With an efficient coupling chemistry identified, we characterized the translation process in greater detail using modified substrates that contain only one functional group

for coupling and therefore terminate oligomerization (**T6** and **T6-f**) (Figure 2.6a and b). T6-f is conjugated to the Alexa Fluor 647 fluorophore (Figure 2.6b) and therefore its incorporation can be readily quantitated. We prepared three templates in which the terminator codon (CATCA), analogous to a stop codon, was placed at either the second, fourth, or sixth codon position. Denaturing PAGE analysis revealed that oligomerization with substrates **6**, **12**, and **T6-f** proceeded until the codon position complementary to the terminator substrate, and then stopped (Figure 2.6c). Fluorescent imaging revealed that T6-f was indeed incorporated in a sequence-programmed manner (Figure 2.6c, right gel, lane 3-5). In contrast, no Alexa Fluor 647 fluorescence was observed in a control reaction containing all substrates but using a template lacking a terminator codon (Figure 2.6c, right gel, lane 2), indicating that the terminator is incorporated only when its corresponding codon is present in the template. Taken together, these results demonstrate that this DNA-templated translation strategy proceeds in a sequence-specific manner. An in-depth study of the sequence specificity of the system is presented below.

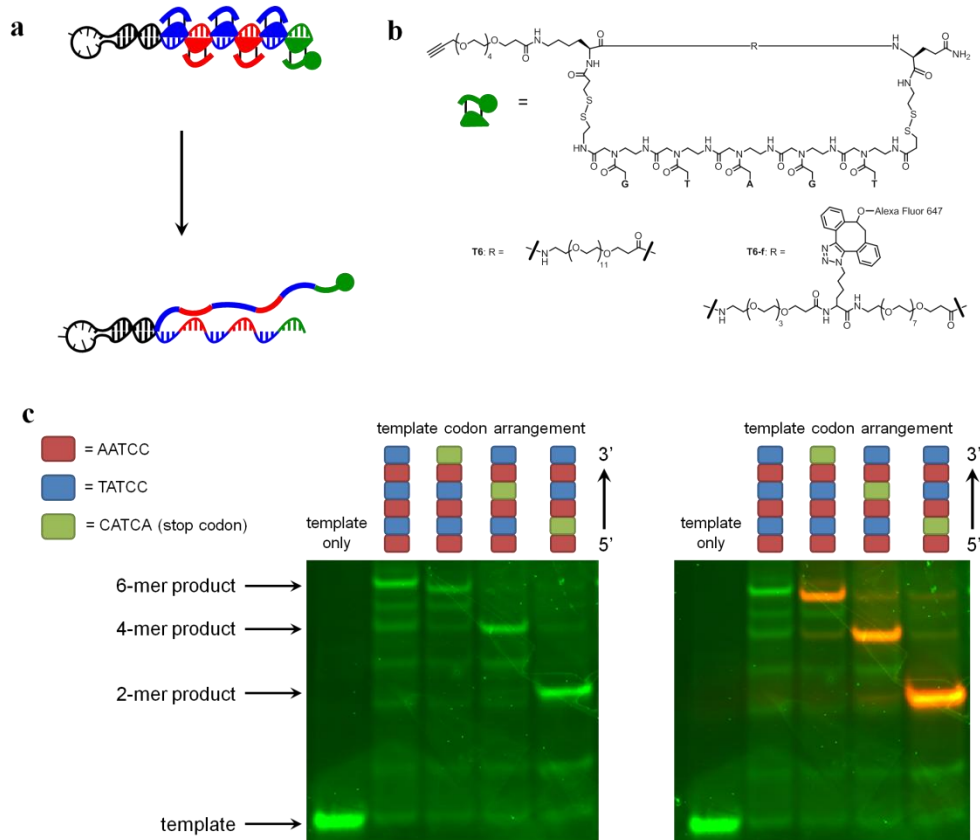


Figure 2.6. (a) A fluorescent ‘terminator’ substrate was used to detect full-length translation products. (b) Structure of terminator substrates **T6-f** and **T6**. (c) Dual-channel fluorescent image of a denaturing PAGE analysing translation reactions containing **T6-f**. Single-stranded DNA was stained with SYBR gold. The coloured rectangles above each lane indicate the arrangement of the codons on the template used in that experiment. The left image shows SYBR gold fluorescence. The right image shows both SYBR gold (green) and Alexa Fluor 647 fluorescence (red).

In order to unambiguously determine the presence and the identity of the full-length product, we analyzed translation reactions using high-resolution electrospray ionization (ESI) mass spectrometry. Products from the DNA-templated oligomerization of **6**, **12**, and a terminator substrate (**T6**) along with a template containing the terminator

codon were subjected to disulfide linker cleavage and analyzed by denaturing PAGE. The largest molecular weight band hypothesized to be the full-length translation product was excised and the material was extracted from the gel. The DNA portion of the product was digested using P1 nuclease and the remaining material was analyzed by ESI-LC-MS (Figure 2.7). The mass spectrum revealed multiply charged species of a single mass consistent with the full-length synthetic polymer (observed mass = 9,980.0 Da; expected mass = 9,979.8 Da) (Figure 2.7). Minor truncation products from the translation that were resolved by gel were also analyzed in the same way as the full-length products using ESI-LC-MS. They were proved to be species that failed to form the triazol linkages under CuAAC conditions (Figure 2.8). These observations collectively support our interpretation of the PAGE data and suggest that the major product generated using this DNA-templated translation strategy is the sequence-programmed full-length synthetic polymer.

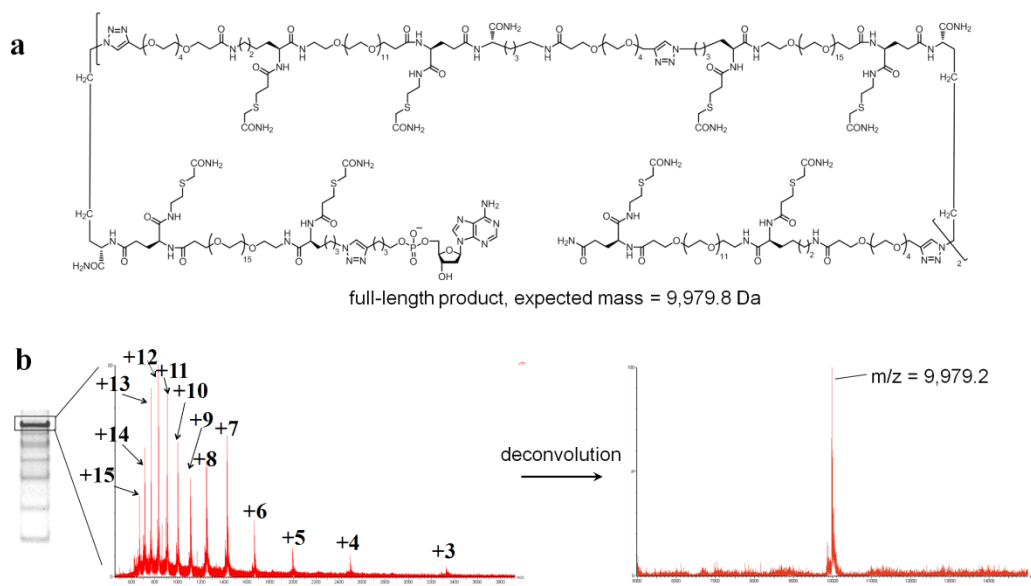


Figure 2.7

Figure 2.7 (Continued)

Figure 2.7. (a) Structure of the full-length product with **T6** incorporated as the terminator after P1 nuclease digestion. (b) MS characterization of the gel-purified full-length product. The left spectrum shows the original multiple charged states of a single molecular species in the ESI analysis, with the numbers indicating the charge of each ion; the right spectrum is deconvoluted from the left spectrum.

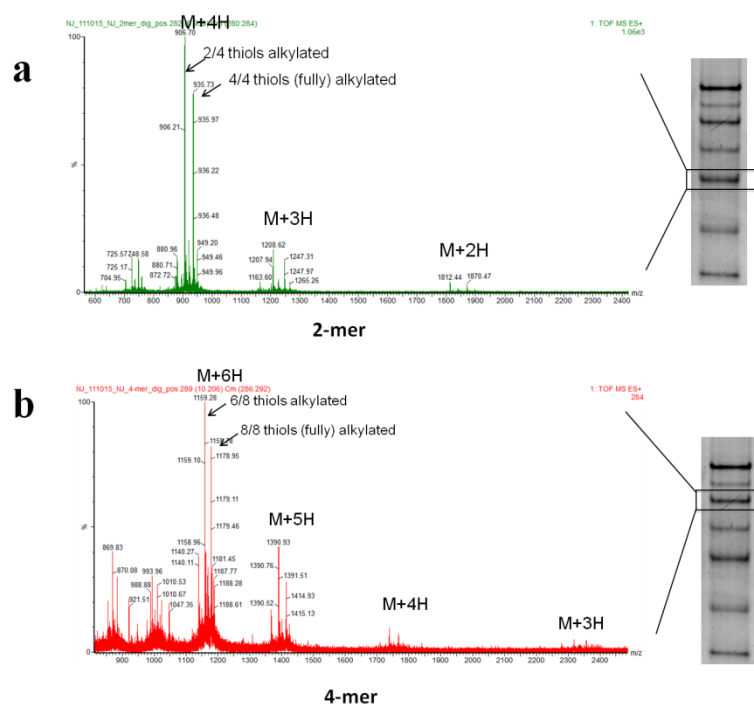


Figure 2.8. ESI-LC-MS analyses of the truncated products. (a) ESI-LC-MS spectrum of the PAGE band hypothesized to be the 2-mer truncated product. There were two major ion species in the spectrum. One corresponds to a product in which all four free thiols generated by disulfide cleavage are alkylated; the other species corresponds to a product in which two out of four thiols are alkylated. (b) ESI-LC-MS spectrum of the PAGE band hypothesized to be the 4-mer truncated product. Once again, two major ion species corresponding to full and partial thiol alkylation. The translation was performed using template **D2K-6f** (containing six codons) and substrates **6**, **12**, and **T6**. No species corresponding to missing thiol alkylation were observed for the full-length product (Figure 2.7b). We speculate that the ion species corresponding to two missing thiol alkylations may arise from a thiol-yne reaction in which two free thiols react with the remaining alkyne groups in truncated products that failed to couple in CuAAC.

2.5 Sequence-Specificity of Translation

The sequence specificity of any synthetic polymer translation strategy is crucial to its suitability for synthetic polymer evolution. To test the sequence specificity of this system in greater depth, and to characterize its compatibility with a template containing multiple codons, we attempted the translation of a DNA template containing six different codons from two codon sets (Figure 2.9). Codons from set 1 encoded the azide building blocks, while codons from set 2 encoded the alkyne building blocks. Codons from set 1 and set 2 alternated along the coding region of the template. Each translation reaction used a mixture of PNA substrates comprising five AA/BB-type azide or alkyne substrates and one terminator substrate, each encoded by a different codon. Only if the terminator substrate is correctly installed opposite its complementary codon, and if the other azide and alkyne bifunctional substrates are incorporated sequence specifically, will oligomerization generate a product of the desired length.

For all six terminator substrates, the predominant translation product was the polymer of expected length (Figure 2.9). These results establish that all 12 substrates tested (six bifunctional substrates and six terminators) containing six different PNA anticodons are incorporated in a template sequence-programmed manner, even in the presence of a stoichiometric excess of non-cognate substrates.

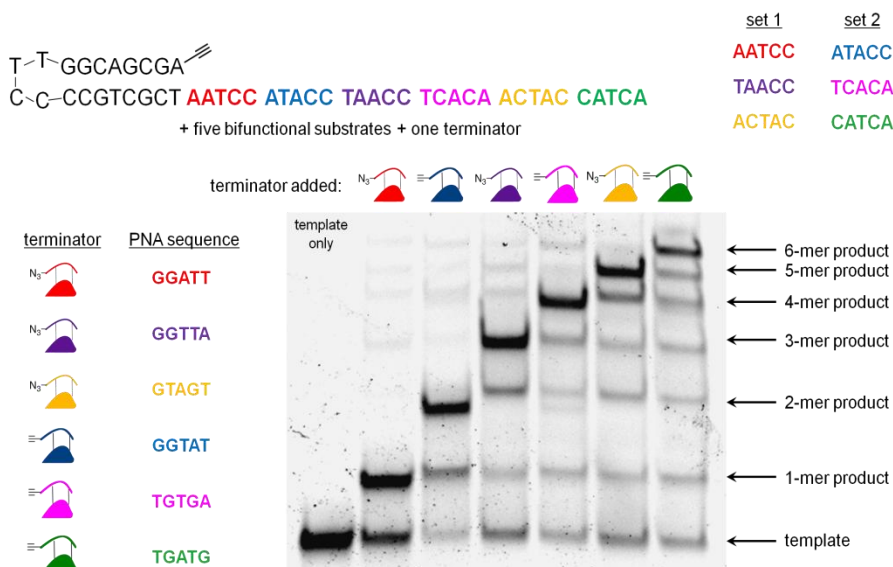


Figure 2.9. Each reaction uses a DNA template contains a coding region with six different codons from two codon sets. Every codon from one set is followed by a codon from the other set. Each translation reaction contained five bifunctional AA/BB-type substrates and one terminator substrate. Because incorporation of the terminator ends oligomerization at the codon position in which it is incorporated, polymer lengths reflect the sequence specificity of the translation process.

2.6 Translation of DNA into Longer, Structurally more Diverse Synthetic Polymers

Since DNA hybridization in this approach is spatially separated from synthetic polymer building blocks, this system in theory should support the translation of DNA templates into a variety of synthetic polymers beyond the PEG-based polymers used in our initial studies. To test this possibility, we designed macrocyclic substrates that contain synthetic polymer building blocks of greater structural diversity. We synthesized macrocyclic substrates containing β -peptide and α -(D)-peptide backbones with a variety of proteinogenic and non-proteinogenic aryl, alkyl, amino, and carboxyl side chains

(Figure 2.10a and c). We performed translation reactions with these structurally diverse building blocks as described above with the terminator substrates programmed to be incorporated at the sixth position of each oligomer.

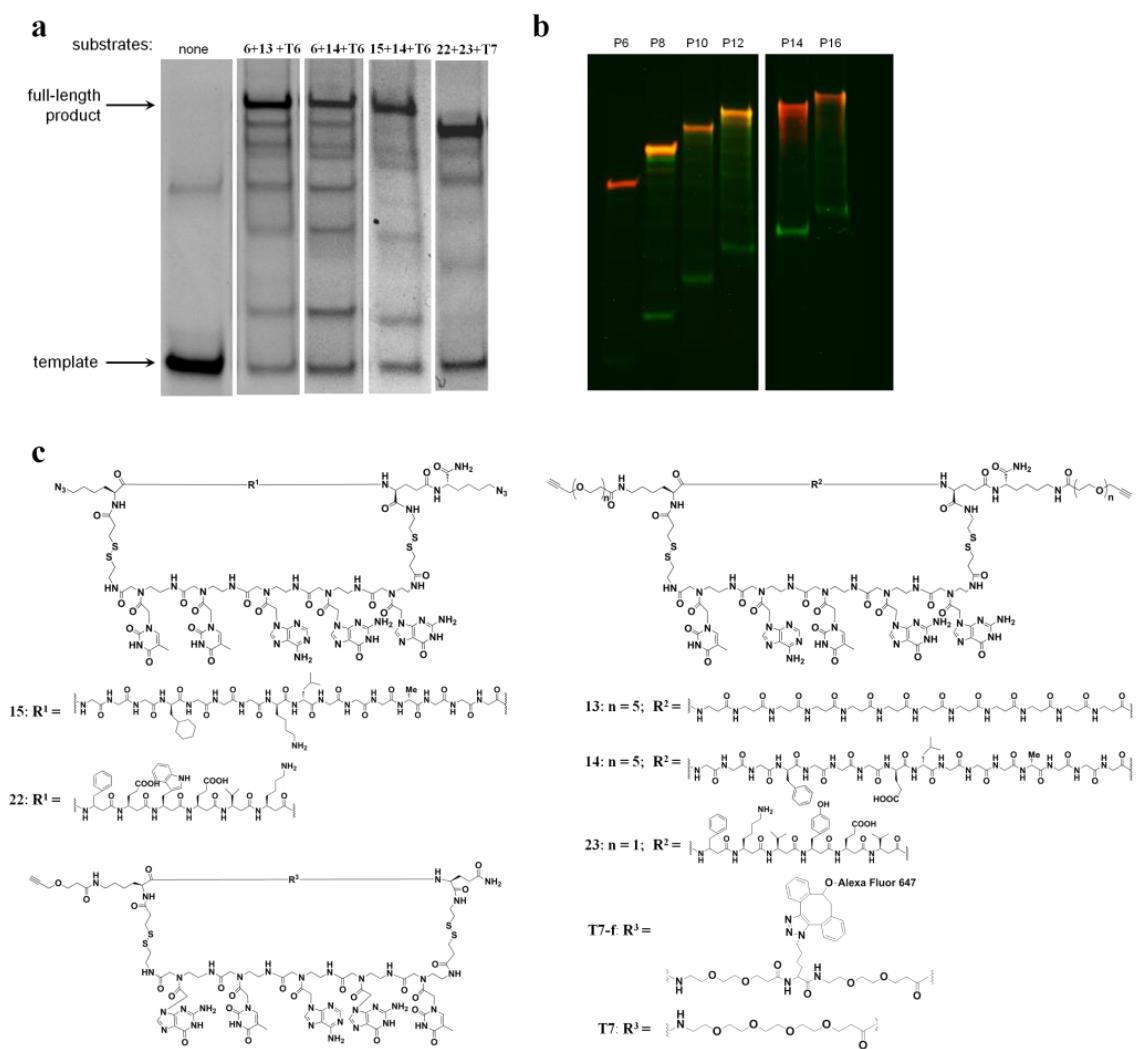


Figure 2.10

Figure 2.10 (Continued)

Figure 2.10. (a) Denaturing PAGE analysis of PEG substrate **6** co-polymerizing with β -peptide substrate **13**, PEG substrate **6** co-oligomerizing with α -(D)-peptide substrate **14**, substrate **14** co-oligomerizing with α -(D)-peptide substrate **15**, and densely functionalized β -peptide substrate **22** co-oligomerizing with densely functionalized β -peptide substrate **23**. The DNA template used in these experiments contained six codons, ending with the 'stop' codon that recruits the terminator substrate **T6** or **T7**. (b) Translation of DNA into longer non-nucleic acid polymers. Translation reactions of DNA templates containing 6, 8, 10, 12, 14 and 16 pentamer codons were performed using substrates **22** and **23**. The last codon of each template encoded the incorporation of the Alexa Fluor 647-linked terminator substrate **T7-f**. The denaturing PAGE was stained with SYBR gold. The green channel shows SYBR gold fluorescence and the red channel shows Alexa Fluor 647 fluorescence. Lanes are marked P6, P8 and so on, reflecting the number of the codons in the template. All lanes contained 10 pmol template for translation reaction. (c) Structures of substrates **13**, **14**, **15**, **22** and **23** and terminators **T7-f** and **T7**.

Denaturing PAGE analysis revealed that **13**, an alkynyl substrate with a β -peptide building block co-oligomerized successfully with azido-PEG substrate **6** in the presence of a DNA template to provide full-length β -peptide-containing hexamer products (Figure 2.10a, lane 2). Similarly, substrate **14**, an alkyne-linked α -(D)-peptide substrate, co-oligomerized successfully with azido-PEG substrate **6** to provide full-length products in comparable yield (Figure 2.10a, lane 3). An azide-linked α -(D)-peptide substrate (**15**) also co-oligomerized with α -(D)-peptide substrate **14** to yield full-length α -(D)-peptide synthetic polymer as the major product (Figure 2.10a, lane 4). Finally, an azide linked, fully functionalized β -peptide substrate (**22**) co-oligomerized with an alkynyl, fully functionalized β -peptide substrate (**23**) to yield full-length products with excellent efficiency (Figure 2.10a, lane 5). A unique feature of **22** and **23** is that they both contain

minimized spacers between the synthetic polymer building blocks and the coupling functional groups to enrich the functionally and structurally meaningful variable region of the synthetic polymer product. ESI-LC-MS analysis after linker cleavage, gel purification, and P1 nuclease digestion confirmed the mass of the full-length products (for **6** + **13**, observed mass = 10,487.0 Da; expected mass = 10,487.2 Da; for **6** + **14**, observed mass = 11,072.0 Da; expected mass = 11,071.6 Da; for **14** + **15**, observed mass = 12,195.0 Da; expected mass = 12,194.6 Da; for **22** + **23**, observed mass = 9,645.3 Da; expected mass = 9643.5 Da) (Figure 2.11). Taken together, these examples demonstrate that the strategy developed in this work can sequence-specifically translate DNA templates into synthetic polymers of uniform length containing a variety of backbone structures. To our knowledge, these results also represent the first enzyme-free translation of nucleic acids into synthetic polymers that have no ability to hybridize to DNA or RNA.

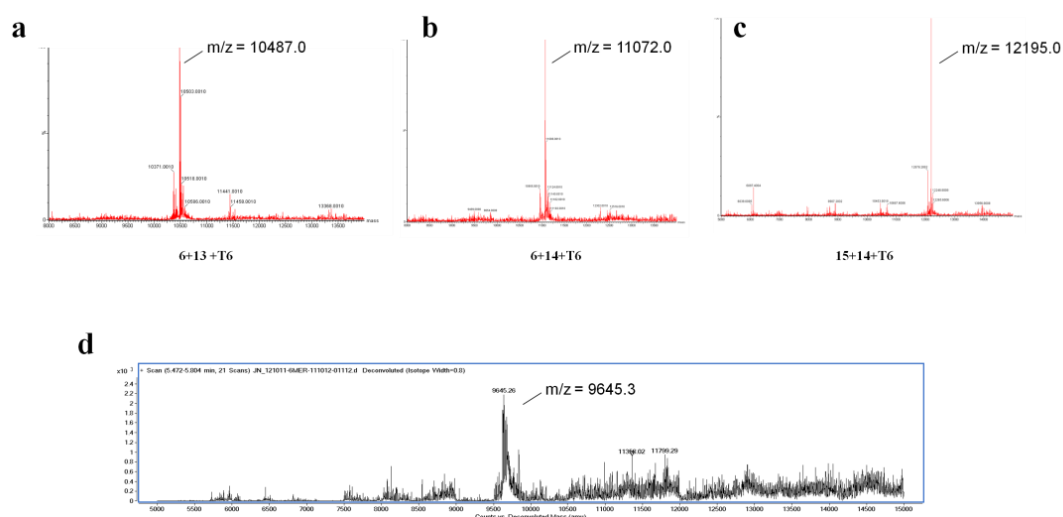


Figure 2.11. (a) Translation using substrates **6**, **13**, and **T6**. (b) Translation using substrates **6**, **14**, and **T6**. (c) Translation using substrates **15**, **14**, and **T6**. (d) Translation using substrates **22**, **23**, and **T7**.

Finally, we characterized the ability of this translation system to generate longer synthetic polymer products. We performed the DNA-templated translation of templates containing six, eight, ten, 12, 14, and 16 pentameric codons using substrates **22**, **23**, and terminator **T7-f**. The fluorescent terminator **T7-f** was structurally similar to **T6-f**, with the exception of shorter synthetic polymer building block and minimized spacer. **T7-f** was included in each translation reaction to enable full-length products to be easily visualized since the last codon in each template uniquely encodes **T7-f** and no other substrates contain fluorophores. Fluorescent imaging of the denaturing PAGE gel revealed that all translation reactions generated full-length polymer product. Translations using longer templates yielded less full-length products, possibly due to more secondary structures formed in the longer template and decrease in total yield as the couplings increase; however, the fraction of the full-length product among total translation product is maintained above 60 %, with the full-length 16-mer synthetic polymer product generated in the longest translation reaction 66 % over all translated products (Figure 2.10b). This 16-mer product, not including the DNA template or PNA adapters, has a molecular weight of 26 kDa and contains 90 β -amino acid residues. These observations indicate the feasibility of translating DNA sequences into sequence-defined synthetic polymer products of molecular weights comparable to those of functional biological polymers.

2.7. Discussion

In this chapter, we designed and implemented a DNA-templated translation system capable of generating sequence-defined synthetic polymers that have no necessary structural similarity to nucleic acids and that do not need to directly hybridize to DNA or RNA. We identified the CuAAC reaction and the AA/BB substrate architecture as key factors to achieve efficient translation. The use of a unique terminator substrate facilitated the analysis of the sequence specificity of the process and also enabled the identification of full-length translation products when coupled with mass spectrometry. This system can be used to generate synthetic polymers containing diverse backbone structures including PEG, β -peptides, and α -(D)-peptides. Building blocks containing different backbones (such as **6+14**) can be co-polymerized sequence-specifically, and full-length products as large as 26 kDa (arising from 16 consecutive substrate couplings and containing 90 β -amino acid residues) were generated using this strategy.

Efficiency and fidelity are essential features of any translation strategy, and in this system arise from at least three design considerations. First, the substrates are organized and subsequently polymerized in a DNA-templated manner. This approach transfers the sequence specificity of Watson-Crick base pairing into the ordering of synthetic polymer building blocks, and also enables reactive groups of adjacent substrates to obtain effective molarities sufficient to drive efficient product formation. In contrast, mixing the substrates in the presence of catalyst but in the absence of a matched DNA template only non-specifically generates dimers, and no higher oligomers are observed (Figure 2.12). Second, the macrocyclic nature of the substrate significantly enhances their reactivity by aligning synthetic polymer building blocks to be coupled in close proximity

on the same side of the DNA double helix. AA/BB architecture of the substrates further minimize undesired reactions including cyclization.¹⁵ Finally, the CuAAC reaction is highly efficient and tolerant of wide variety of functional groups in neutral aqueous solution, and the triazole linkage formed by this reaction is known to be stable under these conditions.^{16,17}

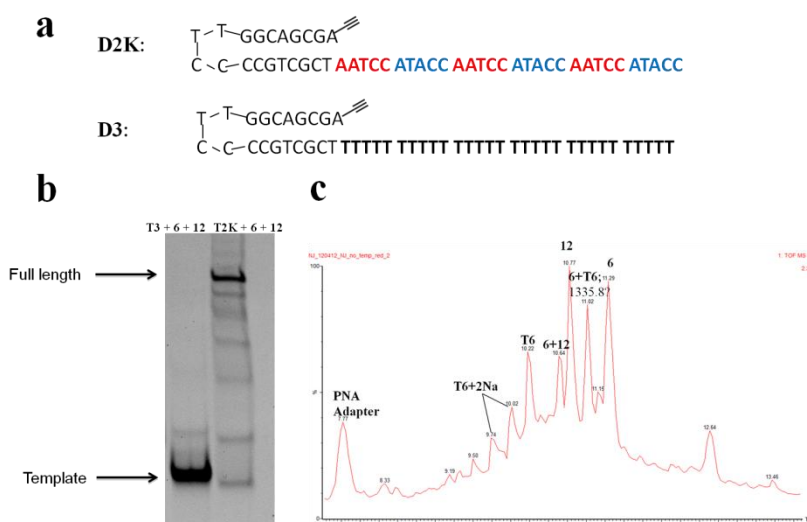


Figure 2.12. Characterization of oligomerization in the absence of a matched template. (a) Oligomerization using a fully mismatched template (**D3**) and a matched template (**D2K**). (b) Non-templated or mismatch-templated products linked to the 5' end of the template would appear as slower migrating bands during PAGE analysis (left lane). The matched-template translation reaction is shown in the right lane. (c) Non-templated reactions among substrates **6**, **12**, and terminator **T6** were analyzed by LC-MS. The ratio of substrates in the reaction mixture was **6:12:T6** = 3:2:1, the same as that of a DNA-templated reaction using **D2K** as the template. The total ion chromatograph (TIC) of the LC after CuAAC and adapter cleavage revealed the presence of only dimers **6+12** and **6+T6**, along with all of the starting monomers; no higher order products were observed. The dimer **6+T6** co-eluted with a species of 1335.8 Da, which was not identified. These results indicate that in the absence of a matched DNA template, coupling between substrates was much less efficient compared to DNA-templated oligomerization. These results suggest that mistranslation from non-specific, non-templated substrate coupling occurs infrequently.

While this translation system does not rely on cells or enzymes and therefore is liberated from the structural requirements associated with ribosomal or polymerase-mediated polymerization, achieving these performance characteristics in the absence of enzymes required that substrates meet a different and substantial set of criteria. The substrates used in this work are macrocyclic, a feature that is necessary to reduce the entropic cost of desired polymerization pathways by limiting conformations unfavorable to reaction (Figure 2.4a, lanes 16 versus lanes 19-21).⁴ Moreover, unlike ribosomal translation in which three-base codon:anticodon hybridization templates the polymerization of single amino acids, this system requires PNA adapters that each contain five nucleobases to enable efficient and sequence-specific hybridization. To ensure hybridization at 4 °C, the T_m of substrates **6** and **12** hybridized to a DNA template is 22.4 °C in the reaction buffer used for polymerization, compared to a $T_m < 10$ °C of a macrocyclic substrate with a PNA adapter sequence containing only four nucleobases.

Sequence-defined polymers in the molecular weight range already achieved by this system include many naturally occurring proteins with remarkable binding and catalytic activities. That said, the extent to which the above requirements limit the functional potential of the resulting synthetic polymers remains to be seen, and it may be necessary to explore additional substrate structures, coupling chemistries, or coding schemes in order to enable the successful evolution of synthetic polymers with desirable binding or catalytic properties. Towards this end, the highly modular nature of the

substrates used in this work facilitates the modification of each substrate component including the PNA adapter, synthetic polymer building block, and linkers.

The translation system developed here has the potential to enable the laboratory evolution of a wide range of synthetic polymers and to reveal the evolutionary potential of macromolecules beyond the reach of previous translation systems. For example, the sequence-specific polymerization of β -peptides that are predisposed to form secondary structures may enable the evolution of β -proteins^{18–20} with novel structures and functions including protease resistance,²¹ improved cell penetration,²² and antibiotic activity.²³

2.8 Methods

General methods

Other than the exceptions noted below, all commercially available reagents and solvents were purchased from Aldrich. Fmoc-protected amino acids and resins were purchased from EMD chemicals. Fmoc-PEG₁₂-OH and Fmoc-PEG₁₆-OH were purchased from Quanta Biodesign. N₃-PEG₄-OH, acetylene-PEG₄-OH, and propargyl-O-propionic acid NHS ester were purchased from Click Chemistry Tools. Reagents for oligonucleotide synthesis were purchased from Glen Research. All PNA monomers were purchased from Polyorg, Inc. All chemical reagents and solvents were used without further purification. The Cu(I) ligand tris(hydroxypropyltriazolyl)amine (THPTA) was synthesized following a previously published protocol.²⁴

Synthesis of the disulfide linker 3-[(2-[(9H-Fluoren-9-ylmethoxy)carbonyl]amino)ethyl)dithio]propanoic acid²⁵

To a solution of 3-mercaptopropanoic acid (4.88 g, 46 mmol) and cysteamine hydrochloride (5.23g, 46 mmol) in water (80 mL) was added DIPEA (12.0 mL, 69 mmol) and one granule (~5 mm diameter) of FeSO₄. The reaction was cooled in an ice bath under continuous stirring for 10 minutes and then was titrated using 30 % hydrogen peroxide. At the end of titration, the color of the mixture turned from pink to pale yellow. The cold solution was acidified with concentrated HCl to pH 2 and stirred for 1 hour. The diacid byproduct was removed by filtration and washed with cold 0.1 M HCl. The aqueous filtrate was extracted by 100 mL EtOAc three times and neutralized with DIPEA. Additional DIPEA (16.0 mL, 92 mmol) was added to the aqueous layer while stirring, followed by 50 mL dioxane and Fmoc-OSu (25.8 g, 77 mmol) in dioxane (50 mL). The reaction was stopped after 5 hours when TLC showed no more new product was formed. The mixture was washed with 50 mL Et₂O and the pH of the aqueous layer was adjusted to 1.0 with concentrated HCl. The whole suspension was extracted with 200 mL EtOAc (50 mL×4) and the combined organic layers were washed with brine, then dried over anhydrous Na₂SO₄. The solvent was evaporated and the crude product was purified by flash column chromatography to yield a white solid (3.0 g, 49 % yield): TLC R_f = 0.50 in 10:1 DCM:methanol; ¹H NMR (500 MHz, CDCl₃) δ 7.75 (d, J = 7.5 Hz, 2H), 7.59 (d, J = 7.5 Hz, 2H), 7.38 (t, J = 7.5 Hz, 2H), 7.30 (t, J = 7.5 Hz, 2H), 5.24 (br, 1H), 4.40 (d, J = 7.1 Hz, 2H), 4.21 (t, J = 7.1 Hz, 1H), 3.51 (m, 2H), 2.90 (t, 2H), 2.79 (m, 4H); ¹³C NMR

(125 MHz, CDCl₃) δ 156.7, 144.0, 141.6, 127.9, 127.3, 125.3, 120.2, 67.0, 47.4, 40.0, 38.1, 34.0, 33.1; MS (ESI⁺): 404.10 (Calculated M-H⁺: 404.09)

Synthesis of Fmoc-L-Lys(propargyl-O-propionyl)-OH (Fmoc-Lys(Pop)-OH)

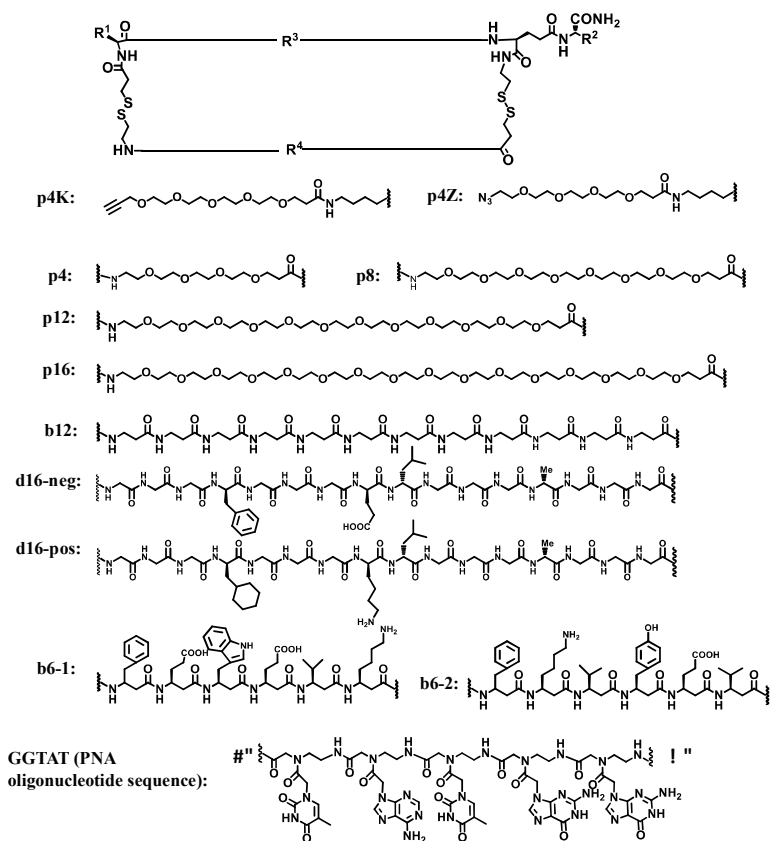
To a solution of Fmoc-L-Lys-OH (3.63 g, 9.9 mmol) in 50 mL DCM was added Propargyl-O-propionic acid NHS ester (2.2 g, 9.9 mmol) and DIPEA (2.0 mL, 11.0 mmol). After 5 hours of reaction, the unreacted Fmoc-Lys-OH was filtered off. The solvent was evaporated *in vacuo* and the remaining residue was dissolved in EtOAc, washed with brine and dried over Na₂SO₄. The crude product was purified on a flash chromatography using eluent DCM: methanol = 20:1 ~5:1 to obtain an off-white solid (2.6 g, 55% yield). TLC R_f = 0.5 in 10:1 DCM: methanol; ¹H NMR (500 MHz, CDCl₃) δ 7.73 (d, J = 7.5 Hz, 2H), 7.57 (d, J = 7.5 Hz, 2H), 7.36 (t, J = 7.5 Hz, 2H), 7.28 (t, J = 7.5 Hz, 2H), 6.57 (br, 1H), 5.93 (br, 1H), 4.41 (br, 1H), 4.34 (d, J = 7.1 Hz, 2H), 4.17 (t, J = 7.1 Hz, 1H), 4.09 (s, 2H), 3.71 (t, J=7.0 Hz, 2H), 3.22 (m, 2H), 2.47 (s, 1H), 2.45 (t, J=7.2 Hz, 2H), 1.85 (m, 1H), 1.73 (m, 1H), 1.48 (m, 2H), 1.40 (m, 2H); ¹³C NMR (125 MHz, CDCl₃) δ 172.5, 156.6, 143.9, 141.5, 127.9, 127.3, 125.4, 120.2, 79.4, 75.5, 67.2, 58.5, 54.1, 47.3, 40.6, 39.3, 36.8, 32.0, 28.9, 22.4; MS (ESI⁺): 479.21 (Calculated M-H⁺: 479.21).

Synthesis of macrocyclic substrates

The synthesis of the linear precursor of **6** was performed on 10 μ mol scale on a Protein Technologies Tribute-UV automated peptide synthesizer using Fmoc chemistry,

with DMF as the solvent and 20% piperidine in DMF as the deprotection reagent. Peptide couplings were performed on Rink amide low-loading resin using 5 eq. Fmoc-protected amino acid activated with 4.5 eq. HATU and 10 eq. *N*-methylmorpholine. The sequence of the linear precursor is NH₂-Lk-*G-G-A-T-T*-Lk-Lys(N₃)-PEG₁₆-Glu(ODmab)-Lys(N₃)-CONH₂, where Lk is the disulfide linker; Lys(N₃) is ϵ -azido lysine; PEG₁₆ is coupled using Fmoc-NH-PEG₁₆-COOH as one residue; Glu(ODmab) is coupled using Fmoc-NH-Glu-ODmab; and italicized letters represent PNA nucleotides. The resulting resin was placed in a glass peptide reaction vessel, and 2 % hydrazine in DMF (2 mL \times 5) was added to deprotect the Dmab group. The resin was washed with DMF (2 mL \times 5), 5 % DIPEA in DMF (2 mL \times 2), and DMF again (2 mL). The linear precursor was cyclized on resin using diisopropylcarbodiimide (DIC, 20 μ L, 125 μ mol) and hydroxybenzotriazole (HOBt, 16.2 mg, 120 μ mol) in 3 mL DMF for 48 h at room temperature. The product was cleaved from the resin by treatment of a mixture (2 mL \times 2) of 95 % trifluoroacetic acid (TFA), 2.5 % water, and 2.5 % triisopropylsilane (TIS). The crude macrocycle was precipitated in diethyl ether and purified by C18 reverse-phase HPLC using 0.1 % TFA in water and 0.1 % TFA in acetonitrile in a linear gradient from 10 % to 40 % acetonitrile as the mobile phase. HPLC fractions were characterized by MALDI mass spectrometry and fractions containing pure macrocycle were combined and lyophilized.

Table 2.1. Structures of Substrates (not including 20 and 21)




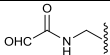
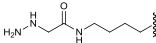
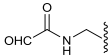
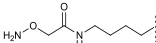
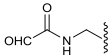
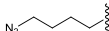

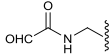
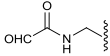
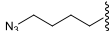
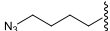
	R ¹	R ²	R ³	R ⁴ (PNA sequence, from N to C ends)
1			p16	AAGGA
2			p16	AAGGA
3			p16	AAGGA
4			p16	AAGGA
5			p16	GGATT
6			p16	GGATT

Table 2.1 (Continued)

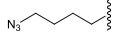
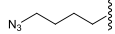
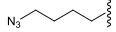

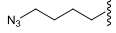
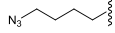
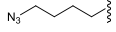

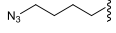
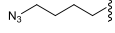


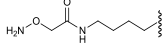
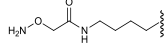
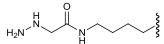
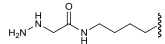
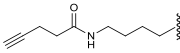
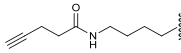
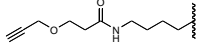
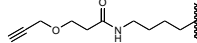
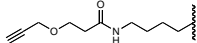
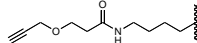
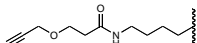
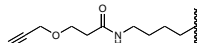
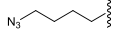
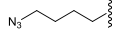
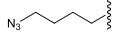
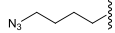
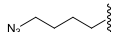
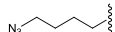
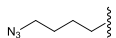
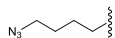
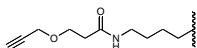
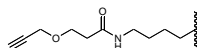
6s			p12	GGATT
6-p0			-	GGATT
6-p4			p4	GGATT
6-p8			p8	GGATT
6-p12			p12	GGATT
7	p4Z	p4Z	p12	GGATT
8			p16	GGTAT
9			p16	GGTAT
10			p16	GGTAT
11			p16	GGTAT
12	p4K	p4K	p12	GGTAT
12s			p8	GGTAT
12-p0-s			-	GGTAT
12-p0-L	p4K	p4K	-	GGTAT
12-p4-s			p4	GGTAT
12-p4-L	p4K	p4K	p4	GGTAT
12-p8-L	p4K	p4K	p8	GGTAT
13	p4K	p4K	b12	GGTAT
14	p4K	p4K	d16-neg	GGTAT
15			d16-pos	GGATT
16			p16	GGTTA
17	p4K	p4K	p12	TGTGA
18			p16	GTAGT

Table 2.1 (Continued)				
19	p4K	p4K	p12	TGATG
22			b6-1	GGATT
23			b6-2	GGTAT

Pre-cleavage modification during substrate synthesis

During the synthesis of **2**, **3**, **9**, **10**, **11**, and **14**, the cyclized substrates were modified before TFA cleavage. For these syntheses Fmoc-Lys(Mmt)-OH was coupled at the modification site in the linear precursor synthesis (Table 2.2). The resin was first treated with a 1:2:7 mixture of AcOH: trifluoroethanol: DCM (2.5 mL \times 4) to remove the Mmt protecting groups on the lysine side chains, followed by base wash with 5 % DIPEA in DMF (2 mL \times 2). The deprotected material was incubated with 0.1 mmol of one of the following carboxylic acids: (Boc)₃-Haa-OH, (Boc)₂-Aoa-OH, pentynoic acid, or acetylene-PEG₄-acid in DCM in the presence of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC-HCl, 19.2 mg, 0.1 mmol), sulfo-*N*-hydroxysuccinimide (sulfo-NHS, 26.1 mg, 0.12 mmol) and DIPEA (87 μ L, 0.5 mmol). After the modification reaction, TFA cleavage was performed as described above.

Post-cleavage modification during substrate synthesis

The synthesis of some substrates required post-cleavage modification (Table 2.2). To prepare **5**, the lyophilized product was re-dissolved in 0.2 mL 0.1 M pH 5.0 NaOAc buffer, followed by the addition of 2 μ L 0.1 M aqueous NaIO₄. The reaction was

incubated at room temperature in the dark for 30 minutes and quenched by addition of 1 μ L 3 M aqueous glycerol. The crude reaction was subjected to reverse-phase HPLC, and the HPLC fractions containing the desired product were collected and lyophilized as described above.

To prepare substrates **7**, **12**, **12-p0-L**, **12-p4-L**, **12-p8-L**, **13**, **17**, **19**, **T2**, **T4**, **T6**, and **T6-f**, the lyophilized product after TFA cleavage was re-dissolved in 0.2 mL 1 \times PBS buffer and 0.01 mmol of acetylene-PEG₄-NHS ester or N₃-PEG₄-NHS ester in 20 μ L DMSO was added. The reaction was incubated for 2 hours at room temperature and quenched with 25 mL 1 M Tris pH 8.0 buffer. The crude reaction was purified by reverse-phase HPLC, and the HPLC fractions containing the desired product were collected and lyophilized.

The synthesis of substrates **1**, **4**, **6**, **6s**, **6-p0**, **6-p4**, **6-p8**, **6-p12**, **8**, **12s**, **12-p0-s**, **12-p4-s**, **15**, **16**, **18**, **22**, **23**, **T1**, **T3**, **T5**, **T7**, and **T7-f** did not require pre-cleavage or post-cleavage modification steps.

Table 2.2. The building blocks used to assemble substrates (not including 22 and 23)

Substrate	AA ¹	AA ²	Synthetic polymer building block	PNA	Pre-cleavage modification	Post-cleavage modification
1	Fmoc-Dpr(Boc-Ser(tBu))-OH	Fmoc-Lys(Boc)-OH	Fmoc-PEG ₁₆ -OH	AAGGA	-	-
2	Fmoc-Dpr(Boc-Ser(tBu))-OH	Fmoc-Lys(Mmt)-OH	Fmoc-PEG ₁₆ -OH	AAGGA	(Boc) ₃ -Haa-OH	-
3	Fmoc-Dpr(Boc-Ser(tBu))-OH	Fmoc-Lys(Mmt)-OH	Fmoc-PEG ₁₆ -OH	AAGGA	(Boc) ₂ -Aoa-OH	-

Table 2.2 (Continued)

4	Fmoc-Bpg-OH	Fmoc-Lys(N3)-OH	Fmoc-PEG ₁₆ -OH	AAGGA	-	-
5	Fmoc-Dpr(Boc-Ser(tBu))-OH	Fmoc-Dpr(Boc-Ser(tBu))-OH	Fmoc-PEG ₁₆ -OH	GGATT	-	NaIO ₄
6	Fmoc-Lys(N3)-OH	Fmoc-Lys(N3)-OH	Fmoc-PEG ₁₆ -OH	GGATT	-	-
6s	Fmoc-Lys(N3)-OH	Fmoc-Lys(N3)-OH	Fmoc-PEG ₁₂ -OH	GGATT	-	-
6-p0	Fmoc-Lys(N3)-OH	Fmoc-Lys(N3)-OH	-	GGATT	-	-
6-p4	Fmoc-Lys(N3)-OH	Fmoc-Lys(N3)-OH	Fmoc-PEG ₄ -OH	GGATT	-	-
6-p8	Fmoc-Lys(N3)-OH	Fmoc-Lys(N3)-OH	Fmoc-PEG ₈ -OH	GGATT	-	-
7	Fmoc-Lys(Boc)-OH	Fmoc-Lys(Boc)-OH	Fmoc-PEG ₁₂ -OH	GGATT	-	N ₃ -PEG ₄ -NHS
8	Fmoc-Lys(Boc)-OH	Fmoc-Lys(Boc)-OH	Fmoc-PEG ₁₆ -OH	GGTAT	-	-
9	Fmoc-Lys(Mmt)-OH	Fmoc-Lys(Mmt)-OH	Fmoc-PEG ₁₆ -OH	GGTAT	(Boc) ₂ -Aoa-OH	-
10	Fmoc-Lys(Mmt)-OH	Fmoc-Lys(Mmt)-OH	Fmoc-PEG ₁₆ -OH	GGTAT	(Boc) ₃ -Haa-OH	-
11	Fmoc-Lys(Mmt)-OH	Fmoc-Lys(Mmt)-OH	Fmoc-PEG ₁₆ -OH	GGTAT	Pentynoic acid	-
12	Fmoc-Lys(Boc)-OH	Fmoc-Lys(Boc)-OH	Fmoc-PEG ₁₂ -OH	GGTAT	-	Acetylene-PEG ₄ -NHS
12s	Fmoc-Lys(Pop)-OH	Fmoc-Lys(Pop)-OH	Fmoc-PEG ₈ -OH	GGTAT	-	-
12-p0-s	Fmoc-Lys(Pop)-OH	Fmoc-Lys(Pop)-OH	-	GGTAT	-	-
12-p0-L	Fmoc-Lys(Boc)-OH	Fmoc-Lys(Boc)-OH	-	GGTAT	-	Acetylene-PEG ₄ -NHS
12-p4-s	Fmoc-Lys(Poe)-OH	Fmoc-Lys(Poe)-OH	Fmoc-PEG ₄ -OH	GGTAT	-	-
12-p4-L	Fmoc-Lys(Boc)-OH	Fmoc-Lys(Boc)-OH	Fmoc-PEG ₄ -OH	GGTAT	-	Acetylene-PEG ₄ -NHS
12-p8-L	Fmoc-Lys(Boc)-OH	Fmoc-Lys(Boc)-OH	Fmoc-PEG ₈ -OH	GGTAT	-	Acetylene-PEG ₄ -NHS
13	Fmoc-Lys(Boc)-OH	Fmoc-Lys(Boc)-OH	(Fmoc-β-Ala-OH) × 12	GGTAT	-	Acetylene-PEG ₄ -NHS

Table 2.2 (Continued)

14	Fmoc-Lys(Mmt)-OH	Fmoc-Lys(Mmt)-OH	(Fmoc-Gly-Gly-Gly-OH) × 4, Fmoc-D-Ala-OH, Fmoc-D-Phe-OH, Fmoc-D-Glu(OtBu)-OH, Fmoc-Leu-OH	GGTAT	Acetylene-PEG ₄ -acid	-
15	Fmoc-Lys(N3)-OH	Fmoc-Lys(N3)-OH	(Fmoc-Gly-Gly-Gly-OH) × 4, Fmoc-D-Ala-OH, Fmoc-D-Phe-OH, Fmoc-D-Lys(Boc)-OH, Fmoc-D-Leu-OH	GGATT	-	-
16	Fmoc-Lys(N3)-OH	Fmoc-Lys(N3)-OH	Fmoc-PEG ₁₆ -OH	GGTTA	-	-
17	Fmoc-Lys(Boc)-OH	Fmoc-Lys(Boc)-OH	Fmoc-PEG ₁₂ -OH	TGTGA	-	Acetylene-PEG ₄ -NHS
18	Fmoc-Lys(N3)-OH	Fmoc-Lys(N3)-OH	Fmoc-PEG ₁₆ -OH	GTAGT	-	-
19	Fmoc-Lys(Boc)-OH	Fmoc-Lys(Boc)-OH	Fmoc-PEG ₁₂ -OH	TGATG	-	Acetylene-PEG ₄ -NHS
22	Fmoc-Lys(N3)-OH	Fmoc-Lys(N3)-OH	Fmoc-β-Lys-OH, Fmoc-β-Val-OH, Fmoc-β-Trp-OH, Fmoc-β-Glu-OH, Fmoc-β-Phe-OH	GGATT	-	-
23	Fmoc-Lys(Pop)-OH	Fmoc-Lys(Pop)-OH	Fmoc-β-Lys-OH, Fmoc-β-Val-OH, Fmoc-β-Tyr-OH, Fmoc-β-Glu-OH, Fmoc-β-Phe-OH	GGTAT	-	-

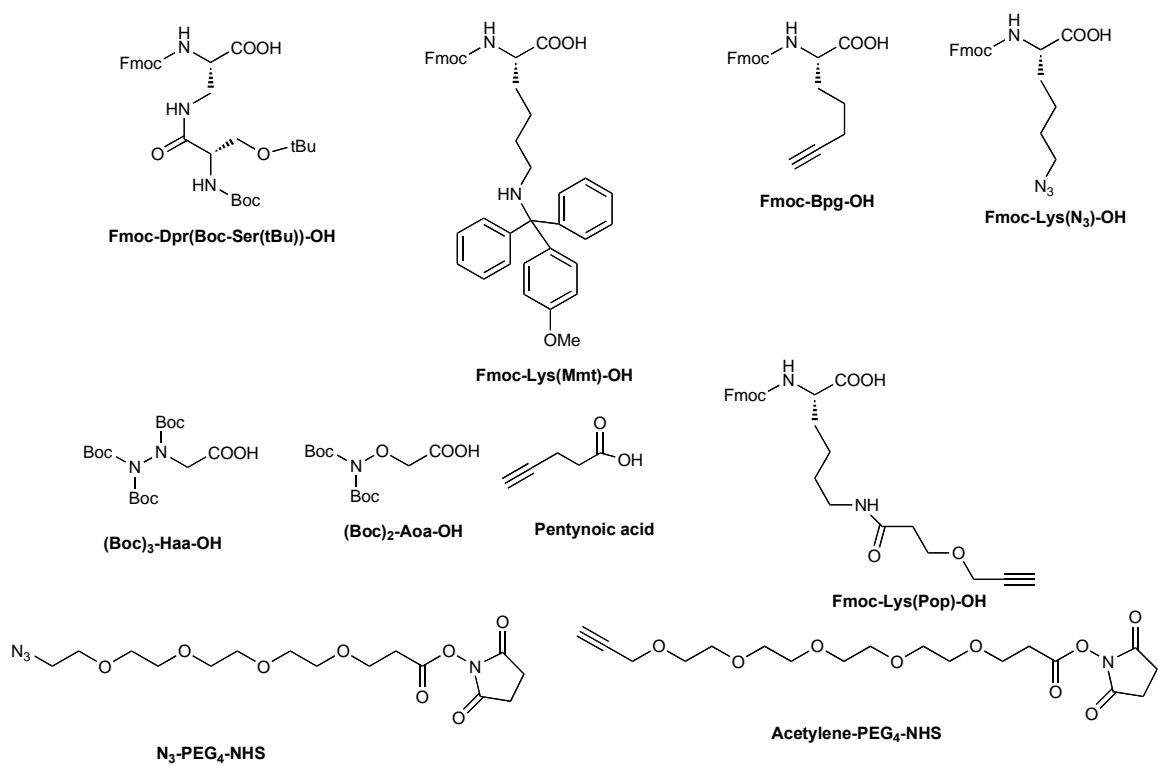


Figure 2.13. Molecular structures for **Table 2.2**

Synthesis of substrates 20 and 21

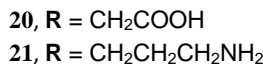


Figure 2.14. Molecular structures of substrates 20 and 21

The syntheses of **20** and **21** start with Fmoc-Glu(Wang resin LL)-ODmab to introduce the C-terminal carboxylate functionality. Fmoc-PEG₁₆-OH is first coupled to the resin as the synthetic polymer building block. The rest of the syntheses of these two substrates are identical to that of **6**. Substrate **20** installed a carboxylate group (from glutamic acid) as the N-terminal functional group for cross-linking, while substrate **21** installed an amino group (from lysine) as the N-terminal functional group.

Synthesis of uncyclized substrates 6-uncyc and 12-uncyc

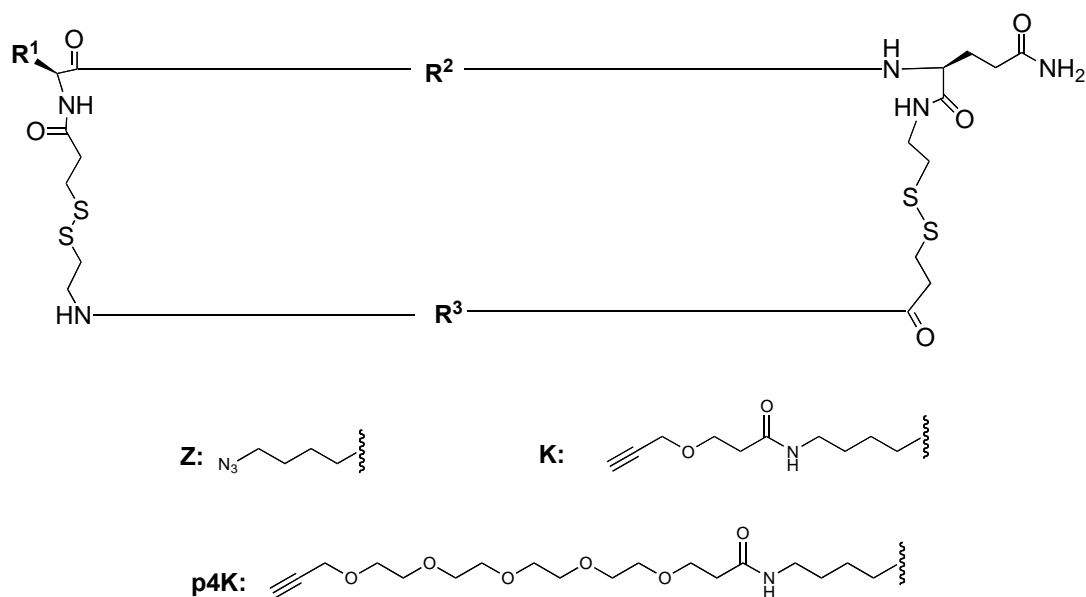
The synthesis of **6-uncyc** is the same as that of **6** except there is no cyclization step before TFA cleavage from the resin. **12-uncyc** was synthesized by replacing the cyclization step with an acetylation step in which the linear precursor of **12** was treated with an acetylation mixture containing 5 % Ac₂O and 6 % 1,2-lutidine in DMF for 5 minutes, followed by washing with 2 mL DMF five times. The rest of the protocol remained the same as that of the synthesis of **12**.

Synthesis of terminator substrates

The terminator substrates were synthesized using modified versions of the syntheses of bifunctional substrates **6** and **12** (Table 2.3). The amino acid reagent used to install the C-terminal functional group (AA¹ in Figure 2.3) was omitted and the linear precursor synthesis directly started from Fmoc-Glu-ODmab. The rest of the synthesis protocol, including post-cleavage modifications, was unchanged.

The synthesis of **T6-f** included the coupling of Fmoc-Lys(N₃)-OH as part of the synthetic polymer building block module, after the coupling of Fmoc-PEG₈-OH and followed by the coupling of Fmoc-PEG₄-OH. After post-cleavage modification, HPLC purification, and lyophilization, the product was re-dissolved in 0.2 mL 1 × TBS buffer and 10 μmol of Alexa Fluor 647 DIBO in DMSO was added. The reaction was shielded from light for 1 hour at room temperature and then purified by reverse-phase HPLC. The HPLC fraction containing the desired product, which exhibited strong UV absorbance at both 254 nm and 650 nm, was collected and lyophilized.

Table 2.3. Terminator substrates



Terminator substrate	R^1	R^2	R^3	Post-cleavage modification
T1	Z	PEG ₁₆	GGATT	-
T2	p4K	PEG ₁₂	GGTAT	Acetylene-PEG ₄ -NHS
T3	Z	PEG ₁₆	GGTTA	-
T4	p4K	PEG ₁₂	TGTGA	Acetylene-PEG ₄ -NHS
T5	Z	PEG ₁₆	GTAGT	-
T6	p4K	PEG ₁₂	TGATG	Acetylene-PEG ₄ -NHS
T6-f	p4K	2 × PEG ₆	TGATG	(i) Acetylene-PEG ₄ -NHS (ii) Alexa Fluor 647 DIBO
T7	Z	PEG ₄	TGATG	-
T7-f	K	2 × PEG ₂	TGATG	i) Acetylene-PEG ₄ -NHS (ii) Alexa Fluor 647 DIBO

Substrate characterization

Mass spectrometry was the primary method used to characterize the substrates. MALDI MS was used to analyze each substrate. Substrates **6**, **8**, **12**, **13**, **14**, and **15** were additionally characterized by ESI-LC-MS to further characterize the compounds. Substrate **8** also underwent disulfide cleavage assay to generate two cleaved fragments for ESI-LC-MS assay to confirm that the cyclization step was successful.

A. MALDI analysis

HPLC fractions that were candidates for containing desired products were directly subjected to MALDI analysis. A 1- μ L sample of the candidate fraction was mixed with 1 μ L saturated sinapic acid solution in 50 % aqueous acetonitrile containing 0.1 % TFA on a MALDI plate. After the spot dried, the sample was subjected to MALDI mass spectrometry using a Waters MALDI micro MX - TOF mass spectrometer. The MALDI spectra of all substrates can be found in Section III of this document.

B. ESI-LC-MS analysis

50 pmol of substrate in 10 μ L 0.1 M ammonium bicarbonate pH 8.0 buffer was injected into a Q-ToF Micro mass spectrometer (Waters) LC-MS system directly. To confirm cyclization, substrate **8** was co-injected with 10 mM TCEP to cleave the disulfide linkers for ESI-LC-MS analysis. The correctly cyclized product showed two cleaved fragments consistent with the expected masses.

Preparation of DNA templates

DNA templates for translation reactions were synthesized with 5' modifications. The 5'-carboxylic acid group was installed using the 5'-carboxy modifier C10 phosphoramidite (Glen Research, Cat No. 10-1935); 5'-alkynyl modification was installed using the 5'-hexynyl phosphoramidite (Glen Research, Cat No. 10-1908); internal Cy3 fluorophore modification was installed using a Cy3TM phosphoramidite (Glen Research, Cat No. 10-5913); 5'-diol modification and the resulting aldehyde modification are described below. The 5'-carboxyl modification and 5'-alkynyl modification were installed according to the manufacturer's instructions. A typical DNA oligonucleotide purification protocol involved: (i) deprotection and cleavage from solid support with 1:1 ammonium hydroxide: methylamine for 15 min at 65 °C; (ii) reverse-phase HPLC purification using a [8% acetonitrile in 0.1 M TEAA, pH 7] to [40% acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45°C.

To install a 5'-diol modification on a DNA template, first a 5'-MMT-amino modifier 5 phosphoramidite (Glen Research Cat No. 10-1905) was used to incorporate a 5'-amino group in the DNA oligonucleotide synthesis. The MMT group deprotection was performed as the last step of automated oligonucleotide synthesis using a DMT-off synthesis program. The CPG beads after a 0.2 μ mol-scale DNA synthesis were washed with acetonitrile (1 mL \times 3) and dried over vacuum. A monobenzyl tartramide (shown in Figure 2.15, 6.5 mg, 20 μ mol), synthesized and used following literature precedent,²⁶ was dissolved in 0.5 mL dry methanol with 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMT-MM, 5.5 mg, 20 μ mol). The pre-activated

tartramide solution was then added to the CPG solid support and reacted for 2 hours. After the reaction the liquid was decanted and the solid support was washed with DMF (1 mL \times 3) and acetonitrile (1 mL \times 3) and dried under vacuum. The modified CPG support then underwent the standard deprotection and purification protocol described above. The fraction containing tartramide-modified oligonucleotide was collected and lyophilized.

The 5'-aldehyde-modified DNA template was obtained by treating the 5'-diol modified DNA with 1 mM NaIO₄ in 0.1 M pH 5.0 NaOAc buffer for 30 minutes at room temperature. After periodate cleavage the reaction products were subjected to another HPLC purification to remove salts and periodate.

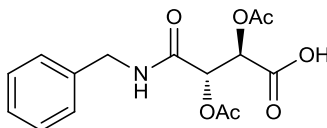


Figure 2.15. (+)-O,O'-diacetyl-L-monobenzyl tartramide

DNA template sequences

Non-standard template functionalities are as follows:

Carboxyl: carboxylic acid C10 modifier.

Diol: monobenzyl tartramide coupled to a Glen Research 5 amino modifier 5.

Hexynyl: hexynyl alkyne modifier.

Aldehyde: diol modification cleaved by NaIO₄.

Cy3: Cy3TM phosphoramidite from Glen Research

BiotinTEG: Biontin TEG modifier was added when ordering from IDT

A. DNA templates in Figure 2.4a

D0C (template for lane 1-2), 5'-Carboxyl-

TTCCGAGCCAGGAATTCGCCCCGGGTCTTCTCCCGGGCGAATTCCTGGCTCGGA
ATCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTT-3'

D1D (template for lane 3-6), 5'-Diol-

AGCGACGGTTCCCCGTCGCTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTT-3'

D1K (template for lane 7-8), 5'-Hexynyl-

AGCGACGGTTCCCCGTCGCTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTT-3'

D3C (template for lane 9-10), 5'-Carboxyl-

AGCGACGGTTCCCCGTCGCTATACCAATCCATACCAATCCATACCAAT CC-3'

D3D (template for lane 11-14), 5'-Aldehyde-

AGCGACGGTTCCCCGTCGCTATACCAATCCATACCAATCCATACCAATCC-3'

D2K (template for lane 15-21), 5'-Hexynyl-

AGCGACGGTTCCCCGTCGCTAATCCATACCAATCCATACCAATCCATA CC-3'

B. DNA templates in Figure 2.4b and 2.5

D2K

C. DNA templates in Figure 2.6c

Lane 1: **D2K**

Lane 2: **D2K**

Lane 3: **D2K-6f**, 5'-Hexynyl-

AGCGACGGTTCCCCGTCGCTAATCCATACCAATCCATACCAATCCCATCA-3'

Lane 4: **D2K-4f**, 5'-Hexynyl-

AGCGACGGTTCCCCGTCGCTAATCCATACCAATCCCATCAAATCCATACC-3'

Lane 5: **D2K-2f**, 5'-Hexynyl-

AGCGACGGTTCCCCGTCGCTAATCCCATCAAATCCATACCAATCCATACC-3'

D. DNA template in Figure 2.10a

D2K-6f

E. DNA templates in Figure 2.10b

P6: **D2K-6f**

P8: **D2K-8f**, 5'-Hexynyl-

AGCGACGGTTCCCCGTCGCTAATCCATACCAATCCATACCAATCCATACCAAT
CCCATCA-3'

P10: **D2K-10f**, 5'-Hexynyl-

AGCGACGGTTCCCCGTCGCTAATCCATACCAATCCATACCAATCCATACCAAT
CCATACCAATCCCATCA-3'

P12: **D2K-12f**, 5'-Hexynyl-

AGCGACGGTTCCCCGTCGCTAATCCATACCAATCCATACCAATCCATACCAAT
CCATACCAATCCATACCAATCCCATCA-3'

P14: **D2K-14f**, 5'-Hexynyl-

AGCGACGGTTCCCCGTCGCTAATCCATACCAATCCATACCAATCCATACCAAT
CCATACCAATCCATACCAATCCATACCAATCCCATCA-3'

P16: **D2K-16f**, 5'-Hexynyl-

AGCGACGGTTCCCCGTCGCTAATCCATACCAATCCATACCAATCCATACCAAT
CCATACCAATCCATACCAATCCATACCAATCCATACCAATCCCATCA-3'

DNA-templated translation

A. Translation for PAGE analysis

Translation reactions were prepared by combining 40 μ L of degassed 0.1 M aqueous HEPES pH 8.0 buffer, 1 μ L of 4 M NaCl, 10 pmol of DNA template, 120 pmole of azide substrate (4 eq. per template codon), 80 pmol of alkyne substrate (4 eq. per template codon), 40 pmol of the terminator substrate, 0.5 μ mol tris(hydroxypropyltriazolyl)amine, THPTA²⁴ and water to a total volume of 50 μ L. The reaction was heated to 95 °C and slowly cooled to 5 °C. The CuAAC reaction was initiated by the addition of 1 μ L 50 mM CuSO₄, followed by the addition of 1 μ L 0.5 M sodium ascorbate. The reaction mixture was maintained at 4 °C overnight. After incubation, the reaction was desalted using a Sephadex minicolumn (Princeton Separations) and analyzed by 10 % denaturing PAGE. The gel was stained with ethidium bromide and analyzed by UV illumination and densitometry.

B. ESI-LC-MS characterization of full-length products

DNA-templated oligomerization was performed by mixing 4 mL of degassed 0.1 M HEPES pH 8.0 buffer, 100 μ L of 4 M NaCl, 1 nmol of DNA template containing six codon positions, 12 nmol of azide substrates, 8 nmol of alkyne substrate, 4 nmol of terminator substrate, and 0.5 mL of 0.1 M Cu(I) ligand THPTA to reach a total volume of 5 mL. The reaction was heated to 95 °C and slowly cooled to 5 °C. CuAAC was initiated by the addition of 100 μ L 50 mM CuSO₄, followed by the addition of 100 μ L 0.5 M sodium ascorbate. The water used to dissolve CuSO₄ and sodium ascorbate was previously deoxygenated by bubbling nitrogen gas through the water for 20 min. The reaction mixture was maintained at 5 °C overnight, then desalted using Nap-10 Sephadex columns. The eluted sample was concentrated under vacuum to 100 μ L, then further desalted using a Princeton Separation minicolumn. The resulted solution was frozen and lyophilized. The sample was dissolved in 60 μ L of water. 10 μ L of 1 M ammonium bicarbonate buffer pH 8 and 2 μ L of 1 M DTT were added to cleave the disulfide linkers and the resulting solution was incubated at 65 °C for 10 min. To alkylate the resulting free thiol groups, 16.8 μ L of 0.5 M iodoacetamide was added, and the resulting solution was incubated in darkness for 30 min at room temperature. The reaction was quenched with 10 μ L of 1 M DTT, followed by incubation in darkness for 30 min at room temperature. The solution was lyophilized and re-dissolved in 20 μ L of 50% aqueous formamide.

The resulting sample was subjected to 10 % denaturing PAGE. The product band was visualized by UV shadowing and excised. The excised gel was homogenized and the translation products were eluted with 400 μ L 10 mM Tris-EDTA pH 7.4 buffer overnight at 37 °C under constant shaking. After filtration to remove the gel particles, the filtrate was concentrated to 100 μ L and desalted using two consecutive Princeton Separation minicolumns. The desalted sample was combined with 190 μ L of 50 mM pH 6.0 NH_4OAc buffer and 5 U of P1 nuclease in 10 μ L pH 6.0 NH_4OAc . The digestion reaction was incubated at 37 °C for 1 hour and lyophilized. The sample was redissolved in 10 μ L 0.1 % TFA and purified by a Ziptip (C-18, Millipore). The sample eluted from the Ziptip was injected into the ESI-LC-MS for analysis.

The ESI-LC-MS spectra in Figure 2.7, Figure 2.8, and Figure 2.11 were all collected and deconvoluted on a Q-ToF Premier LC-mass spectrometer (Waters) except Figure 2.11(d), which was collected and deconvoluted on an Agilent TOF LC-MS system.

C. Mixed codon DNA-templated translation to assess sequence specificity

The oligomerization reaction was prepared by mixing 40 μ L of degassed 0.1 M HEPES pH 8.0 buffer, 1 μ L of 4 M NaCl, 10 pmol of DNA template containing six different codons, and 40 pmol each of five bifunctional substrates and one terminator substrate. Each substrate contained one PNA anticodon complementary to one position on the DNA template. 5 μ L of 0.1 M Cu(I) ligand THPTA, and water were added to a total volume of 50 μ L. The reaction was heated to 95 °C and slowly cooled to 5 °C. CuAAC was initiated by the addition of 1 μ L 50 mM CuSO_4 , followed by the addition of

1 μ L 0.5 M sodium ascorbate. The water used to dissolve CuSO_4 and sodium ascorbate was previously degassed by bubbling nitrogen through the water for 20 minutes. The reaction mixture was maintained at 5 $^{\circ}\text{C}$ overnight. After incubation, the reaction was desalted using a Princeton Separation Sephadex minicolumn and analyzed by 10 % denaturing PAGE. The resulting gel was stained by ethidium bromide and imaged by UV illumination and densitometry.

D. Aldehyde unmasking

Template **D1D** and substrates **1**, **2**, and **3** require unmasking of an aldehyde group before oligomerization can occur. This unmasking was performed immediately before the oligomerization reaction. 10 pmol of DNA template and 240 pmol of substrates were heated to 95 $^{\circ}\text{C}$ and slowly cooled to room temperature in 10 μ L 0.1 M NaOAc pH 5.5 buffer. To initiate aldehyde unmasking, 1 μ L 10 mM NaIO_4 was added. The reaction was incubated at room temperature for 30 minutes, and then desalted using a Princeton Separation minicolumn. The desalted material was used directly in translation reactions.

E. Acylation chemistry

10 pmol of DNA template and 240 pmol of total substrates were used for a single translation reaction. The template and substrates were combined with 40 μ L of 0.1 M MOPS pH 7.0 buffer, 1 μ L of 4 M NaCl, and water to reach a total volume of 50 μ L. The reaction was heated to 95 $^{\circ}\text{C}$ and slowly cooled to 5 $^{\circ}\text{C}$. The acylation reaction was initiated by the addition of 1 μ L 0.5 M aqueous DMT-MM. The reaction mixture was

maintained at 5 °C overnight. After incubation, the reaction was desalted using a Princeton Separation Sephadex minicolumn and separated by 10 % denaturing PAGE-urea gel. The gel was stained by ethidium bromide and imaged by UV illumination and densitometry.

F. Reductive amination

10 pmol of DNA template and 240 pmol of total substrates were used for a single translation reaction. The template and substrates were combined with 40 µL of 0.2 M sodium phosphate pH 8.0 buffer, 1 µL of 4 M NaCl, and water to reach a total volume of 50 µL. The reaction was heated to 95 °C and slowly cooled to 5 °C, unless this annealing step was already performed in the aldehyde unmasking procedure. The reductive amination reaction was initiated by the addition of 1 µL 4 M aqueous NaBH₃CN. The reaction mixture was maintained at 5 °C overnight. After incubation, the reaction was desalted using a Princeton Separation Sephadex minicolumn and separated by 10 % denaturing PAGE-urea gel. The gel was stained by ethidium bromide and imaged by UV illumination and densitometry.

G. Oxime and hydrazone formation

10 pmol of DNA template and 240 pmol of total substrates were used for a single translation reaction. The template and substrates were combined with 15 µL of 0.1 M sodium phosphate pH 7.0 buffer, 1 µL of 4 M NaCl, and water to reach a total volume of 20 µL. The reaction was heated to 95 °C and slowly cooled to 5 °C, unless this annealing step was done in the aldehyde unmasking procedure. The oxime or hydrazone

formation reaction was initiated by the addition of 40 μL 0.1 M aniline in 0.1 M NaOAc pH 5.5 buffer. The reaction mixture was incubated at 5 $^{\circ}\text{C}$ overnight. For hydrazone formation, 1 μL 4 M aqueous NaBH_3CN was then added and the reaction was further incubated at 5 $^{\circ}\text{C}$ for 1 hour. The reaction was desalted using a Princeton Separation Sephadex minicolumn and separated by 10 % denaturing PAGE-urea gel. The gel was stained with ethidium bromide and imaged by UV illumination and densitometry.

References:

1. Rosenbaum, D. M. & Liu, D. R. Efficient and sequence-specific DNA-templated polymerization of peptide nucleic acid aldehydes. *J. Am. Chem. Soc.* **125**, 13924–5 (2003).
2. Kleiner, R. E., Brudno, Y., Birnbaum, M. E. & Liu, D. R. DNA-templated polymerization of side-chain-functionalized peptide nucleic acid aldehydes. *J. Am. Chem. Soc.* **130**, 4646–59 (2008).
3. Brudno, Y., Birnbaum, M. E., Kleiner, R. E. & Liu, D. R. An in vitro translation, selection and amplification system for peptide nucleic acids. *Nat. Chem. Biol.* **6**, 148–55 (2010).
4. Driggers, E. M., Hale, S. P., Lee, J. & Terrett, N. K. The exploration of macrocycles for drug discovery--an underexploited structural class. *Nat. Rev. Drug Discov.* **7**, 608–24 (2008).
5. Hud, N. V, Jain, S. S., Li, X. & Lynn, D. G. Addressing the problems of base pairing and strand cyclization in template-directed synthesis: a case for the utility and necessity of “molecular midwives” and reversible backbone linkages for the origin of proto-RNA. *Chem. Biodiv.* **4**, 768–83 (2007).
6. Winnik, A. End-to-End Cyclization of Polymer Chains. *Acc. Chem. Res.* **18**, 73–79 (1985).
7. Li, X. & Liu, D. R. DNA-templated organic synthesis: nature’s strategy for controlling chemical reactivity applied to synthetic molecules. *Angew. Chem. Int. Ed.* **43**, 4848–70 (2004).
8. Cudic, M. Convenient synthesis of a head-to-tail cyclic peptide containing an expanded ring. *Tetrahedron Lett.* **41**, 4527–4531 (2000).
9. Schmidt, J. G., Christensen, L., Nielsen, P. E. & Orgel, L. E. Information transfer from DNA to peptide nucleic acids by template-directed syntheses. *Nucleic Acids Res.* **25**, 4792–6 (1997).
10. Boehler, C., Nielsen, P. E., Orgel, L. E. & Böhrer, C. Template switching between PNA and RNA oligonucleotides. *Nature* **376**, 578–581 (1995).
11. Goodwin, J. T. & Lynn, D. G. Template-directed synthesis: use of a reversible reaction. *Chem. Commun.* 9197–8 (1992).

12. Cho, B. R., Cho, N. S., Song, K. S., Son, K. N. & Kim, Y. K. Elimination reactions of (E) -2 , 4-dinitrobenzaldehyde O-aryloximes promoted by RO - / ROH buffers in EtOH. *J. Org. Chem.* **2**, 3006–9 (1998).
13. Peng, X., Li, H. & Seidman, M. A Template-Mediated Click-Click Reaction: PNA-DNA, PNA-PNA (or Peptide) Ligation, and Single Nucleotide Discrimination. *Eur. J. Org. Chem.* **2010**, 4194–4197 (2010).
14. Kumar, R. *et al.* Template-directed oligonucleotide strand ligation, covalent intramolecular DNA circularization and catenation using click chemistry. *J. Am. Chem. Soc.* **129**, 6859–64 (2007).
15. Andersen, C. S., Yan, H. & Gothelf, K. V. Bridging one helical turn in double-stranded DNA by templated dimerization of molecular rods. *Angew. Chem. Int. Ed.* **47**, 5569–72 (2008).
16. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click chemistry: diverse chemical function from a few Good reactions. *Angew. Chem. Int. Ed.* **40**, 2004–2021 (2001).
17. El-Sagheer, A. H. & Brown, T. Click chemistry with DNA. *Chem. Soc. Rev.* **39**, 1388–405 (2010).
18. Petersson, E. J. & Schepartz, A. Toward beta-amino acid proteins: design, synthesis, and characterization of a fifteen kilodalton beta-peptide tetramer. *J. Am. Chem. Soc.* **130**, 821–3 (2008).
19. Goodman, C. M., Choi, S., Shandler, S. & DeGrado, W. F. Foldamers as versatile frameworks for the design and evolution of function. *Nat. Chem. Biol.* **3**, 252–62 (2007).
20. Martinek, T. a. & Fulop, F. Side-chain control of beta-peptide secondary structures. Design principles. *Eur. J. Biochem.* **270**, 3657–3666 (2003).
21. Frackenpohl, J., Arvidsson, P. I., Schreiber, J. V & Seebach, D. The outstanding biological stability of beta- and gamma-peptides toward proteolytic enzymes: an in vitro investigation with fifteen peptidases. *Chembiochem* **2**, 445–55 (2001).
22. Bautista, A. D., Appelbaum, J. S., Craig, C. J., Michel, J. & Schepartz, A. Bridged beta(3)-peptide inhibitors of p53-hDM2 complexation: correlation between affinity and cell permeability. *J. Am. Chem. Soc.* **132**, 2904–6 (2010).
23. Olsen, C. *et al.* Antimicrobial, hemolytic, and cytotoxic activities of beta-peptoid-peptide hybrid oligomers: improved properties compared to natural AMPs. *Chembiochem* **11**, 1356–60 (2010).

24. Chan, T. R., Hilgraf, R., Sharpless, K. B. & Fokin, V. V. Polytriazoles as copper(I)-stabilizing ligands in catalysis. *Org. Lett.* **6**, 2853–5 (2004).
25. Sun, Q., Cai, S. & Peterson, B. R. Selective disruption of early/recycling endosomes: release of disulfide-linked cargo mediated by a N-alkyl-3beta-cholesterylamine-capped peptide. *J. Am. Chem. Soc.* **130**, 10064–5 (2008).
26. Gartner, Z. J. *et al.* DNA-templated organic synthesis and selection of a library of macrocycles. *Science* (80-.). **305**, 1601–5 (2004).

Chapter Three

Developments towards an *In Vitro* Selection System of Functional Synthetic Polymers

Jia Niu, Zhen Chen, John Guilinger, Lynn McGregor, and David R. Liu

Zhen Chen helped to optimize the reaction condition for library translation and primer extension reaction. John Guilinger wrote the MATLAB program to search for DNA codons of the genetic code of the translation system. Lynn McGregor wrote the MATLAB program to analyze the high-throughput sequencing results.

3.1 Introduction

In Chapter Two, enzyme-free translation of DNA into sequence-defined non-nucleic acid synthetic polymers is described. However, most of the examples of translation were performed on a single template with repeating DNA codons. In order for a translation system to generate numerous numbers of synthetic polymers as a library for *in vitro* selection, its ability to maintain efficient and unbiased translation for most of, if not all, templates from a DNA template library is highly desired. In Chapter Three, we seek to design a DNA template codon set (also referred as the “genetic code” in this chapter) and a set of synthetic polymer building blocks to maximize the translation efficiency in a library setting and avoid significant bias during the translation reaction. Furthermore, such designs should also preclude hybridization between template codons and non-cognate building blocks that can reduce the sequence-specificity of translation. Last but not least, the genetic code and building block design should allow the DNA templates and substrates readily prepared in solid phase synthesis by using commercially available or easily synthesized building blocks.

To prepare the synthetic polymer library for *in vitro* selection, several key procedures need to be undertaken. First, since the translation yield of the desired full-length products is hardly perfect, a purification strategy to remove all the non-translated templates and truncated translation products is required for optimal selection results. Second, activities emerging during evolution that are dependent on a particular folded structure of the DNA template must also be suppressed to avoid evolving DNA templates,

rather than synthetic polymers, that contribute to the desired activities. Towards this goal, we will develop a post-translation primer extension strategy to synthesize a complementary DNA strand over the DNA template. Finally, the ability of the translation system to be integrated in the iterative cycles of translation, selection, template-regeneration, and re-translation has to be tested before *in vitro* selection can be performed to evolve functional synthetic polymers.

3.2 Genetic Code Design and the Synthesis of DNA Template Library

The previous study on DNA-templated PNA polymerization by Brudno and Liu has already established a set of fundamental criteria for the genetic code of an enzyme-free translation system to achieve successful library polymerization.¹ These criteria – identical GC/AT and pyrimidine/purine ratios, at least one non-G:T mismatch or two wobble mismatches between any two codons, and no out-of-frame annealing – were experimentally tested and have remained an important reference for the genetic code design of the current work. In addition, specific to the translation system developed in Chapter Two there may be more requirements that the potential genetic code has to meet. First, since the translation system uses two types of alternating codons to encode the incorporation of azido and alkynyl substrates, the genetic code needs to have two sets of codons, with codons of one set, referred here as “male” codons, to encode the incorporation of azido substrates and codons of another set, referred here as “female” codons, to encode the incorporation of alkynyl substrates. The arrangement of male and female codons on the DNA template coding region is such that male codons will only

occupy every other codon positions of the template, with female codons in between. This arrangement can ensure that azide functional groups are adjacent to alkyne functional groups in DNA-templated CuAAC polymerization. Second, since the enzyme-free translation system relies on no other interactions between template codons and substrates than the Watson-Crick hybridization, the DNA template assembled from the designed codons must remain largely single-stranded and accessible to substrate binding. Too much secondary structures within the template coding region are likely to compete with the substrate binding and inhibit the translation. Therefore, to maximize translation yield and avoid significant loss library diversity due to non-translatable templates, secondary structures in the coding region of the DNA template are to be minimized. Third, the designed codons should provide sufficient affinity between the DNA template codon and the substrates in aqueous solution under moderate reaction conditions, as measured by the melting temperature of the DNA-PNA hybridized duplex. A melting temperature too low would render the substrate binding too weak, causing reduced translation efficiency; a melting temperature too high is likely to provide excessive binding energy for the mismatched pairs, causing reduced sequence-specificity. Furthermore, codons with melting temperatures too high and too low likely comprise high purine and pyrimidine contents, respectively, resulting undesired low diversity in codons. Finally, the DNA template library that meets all the criteria above has to be synthetically accessible. Ideally the design of genetic code can enable the DNA template library to be synthesized in a single column on a DNA synthesizer, without the need to perform laborious split and pool synthesis. The split and pool oligonucleotide synthesis of a library containing n

possible DNA codons, with each template in the library having m occurrences of codon positions, would require each cycle of a m -cycle syntheses to be splitted into n separate reactions, totaling $m \times n$ codon syntheses. As m and n grows bigger for larger library size, the split and pool strategy quickly becomes impractical and prohibitively low-yielding due to too large numbers ($m \times n$) of codon syntheses.

To summarize, the criteria for a genetic code design would be:

- DNA templates are comprised of alternating “male” and “female” pentameric codons;
- the sequences of codons differ at least two bases from one another;
- codons have same GC content and pyrimidine content;
- codons must avoid out-of-frame annealing;
- codons provide sufficient and non-excessive affinity between DNA template and the substrates;
- the DNA templates assembled from codons must have minimal secondary structures;
- the DNA template library assembled from codons can be prepared in a single reaction column in one DNA synthesis attempt with commercially available or easily synthesized materials.

To arrive at a synthetic strategy that meets all the above criteria, we speculate that the DNA library synthesis can be accomplished by coupling degenerative oligonucleotide phosphoramidites in a solid phase synthesis. The translation system

requires pentameric DNA codon, but there are no degenerative pentanucleotide phosphoramidites commercially available, while synthesizing them may require much additional efforts and specialized apparatus. To overcome this discrepancy, we attempt to divide a pentameric codon into two coupling steps using a dinucleotide and a trinucleotide, or three coupling steps using two dinucleotides and a mononucleotide. Further investigation on the commercial availability of the trinucleotide phosphoramidites reveals that only 20 out of 64 total possible trimer sequences are commercially available (Glen Research, Table 3.1), potentially limiting the actual numbers of pentameric codons that can be made by combining trimers and dimers. This situation is likely due to the fact that degenerative trinucleotide phosphoramidites are mainly used to synthesize genes for protein expression, and minimally only 20 out of all 64 trimer combinations are needed to encode 20 proteinogenic amino acids. In contrast, all possible dinucleotide phosphoramidites are commercially available (Chemgene, Table 3.2), providing synthetic accessibility to pentameric codons using two dinucleotides plus one mononucleotide in the codon assembly.

Table 3.1. Commercially available trinucleotide phosphoramidites

Sequence	GC fraction	Pyrimidine fraction
AAA	0	0
GAA	1/3	0
AAC	1/3	1/3
ATG	1/3	1/3
CAG	2/3	1/3
GAC	2/3	1/3
GGT	2/3	1/3
TGG	2/3	1/3
ACT	1/3	2/3
ATC	1/3	2/3
CAT	1/3	2/3
GTT	1/3	2/3
TAC	1/3	2/3
CGT	2/3	2/3
CTG	2/3	2/3
GCT	2/3	2/3
TGC	2/3	2/3
CCG	1	2/3
TCT	1/3	1
TTC	1/3	1

Table 3.2. Commercially available dinucleotide phosphoramidites

Sequence	GC fraction	Pyrimidine fraction
AA	0	0
AG	1/2	0
GA	1/2	0
GG	1	0
AT	0	1/2
TA	0	1/2
AC	1/2	1/2
CA	1/2	1/2
GT	1/2	1/2
TG	1/2	1/2
CG	1	1/2
GC	1	1/2
TT	0	1
CT	1/2	1
TC	1/2	1
CC	1	1

To simplify the search of a codon set that satisfy all the criteria of genetic code design and is synthetically accessible using commercial available dinucleotide and trinucleotide phosphoramidites, we decide to arbitrarily fix one criterion and screen for the optimal values of the other variables. In Chapter Two, we demonstrate the successful translation of a DNA template containing two codons: AATCC and ATACC. Each of these two codons has an identical GC fraction of $2/5$, and identical pyrimidine fraction of $3/5$. The melting temperature of a DNA template containing six of each of these two

codons with its matched substrates was determined experimentally as 22.4 °C, sufficient to allow substrates to hybridize with the template with high sequence-specificity at reaction temperatures <17 °C. With these preliminary results, we decide to fix the GC fraction of 2/5 and pyrimidine fraction of 3/5 in our search for the codon set for the construction of DNA template library.

We sought to write a computer program based on MATLAB programming language to search for the possible codons that satisfy all the criteria of the genetic code design (see Methods section for the full MATLAB programs). The program asks for inputs of degenerative dinucleotides and trinucleotides, and can generate all possible pentameric-codon combinations using these dinucleotides and trinucleotides. The program is able to find possible out-of-frame annealing patterns by align two candidate pentameric codons and test if a third codon has a significant overlap (equal or larger than four-base identical sequence) with any part of this ten-nucleotide sequence. The occurrences of significant overlaps are counted and the codon set with the minimal overlaps is displayed. This function of the program can help to narrow down the search and find the candidate sets of codons, each set comprising one subset of male codons and one subset of female codons, to meet the no-out-of-frame-annealing criterion. Subsequently, by interfacing with a Unifold module that computes the folding energy of the single-stranded DNA secondary structures, the MATLAB program can also output the secondary-structure folding energies of all DNA templates of the given length that are made of the candidate codon sets determined by the last step, and display the distribution of these secondary-structure energies.

After searching by the MATLAB program, we were able to obtain a candidate codon set that meets all the criteria for the genetic code design. No out-of-frame annealing was found among the codons within this set, including the potential out-of-frame annealing containing four normal base pairs and one G:T wobble base pair. The codon set is comprised of a male subset and a female subset, with each of the subset containing eight codons. Both the male subset and the female subset are made of two degenerative dinucleotide mixtures and a mononucleotide, represented as follows:

Male codons

$$\begin{pmatrix} TC \\ CT \end{pmatrix} \begin{pmatrix} AC \\ CA \\ TG \\ GT \end{pmatrix} A$$

Female codons

$$A \begin{pmatrix} AC \\ CA \\ TG \\ GT \end{pmatrix} \begin{pmatrix} TC \\ CT \end{pmatrix}$$

where the dinucleotides in the parentheses represent their equal-molar degenerative mixture.

The secondary-structure energies of the 12-codon DNA template made of these male and female codons are also significantly lower than most of the candidate sets, with above 95% of all possible DNA templates having the secondary-structure energies below -5.0 kCal/mol under the given folding conditions (0.2 M monovalent ion concentration,

25 °C) (Figure 3.1). This result indicates the majority of the DNA templates made of this codon set are not well-folded, allowing the accessibility of substrates during translation reaction. Importantly, this codon set can be conveniently assembled by two degenerative dinucleotide mixtures of (CT, TC) and (AC, CA, TG, GT), and an adenosine, allowing the DNA-template library made of this codon set to be synthesized in one reaction column on a DNA synthesizer. Finally, for a potential DNA library comprised of 12-codon DNAs assembled by the codons of this set, the theoretical library diversity would be $8^6 \times 8^6 = 7 \times 10^{10}$.

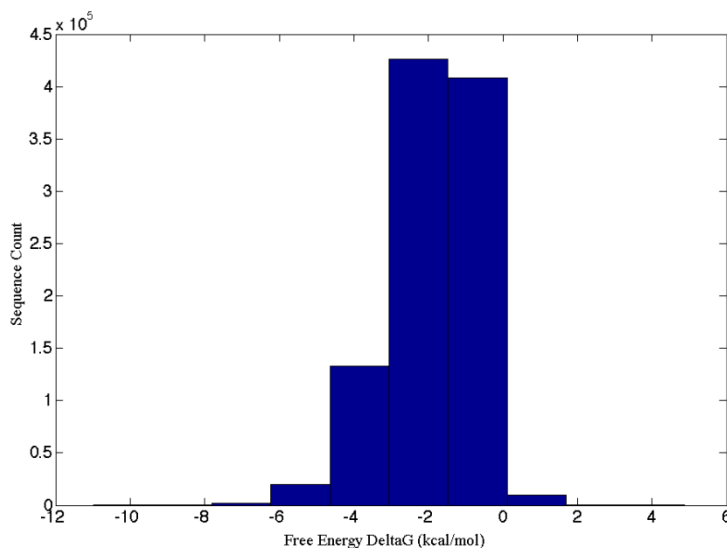


Figure 3.1. Distribution of the secondary-structure folding energies of the DNA templates made of the codon set (CT, TC) (AC, CA, TG, GT)A, and A(AC, CA, TG, GT)(CT, TC).

Having determined the synthetic polymer genetic code, we had the DNA library synthesized by Integrated DNA Technologies. The final sequence of the DNA template library, named as **LibX**, is shown in Figure 3.2. In addition to the coding region where

the variable DNA codons were incorporated by the strategic incorporations of degenerative-dinucleotide phosphoramidite mixtures and an adenosine phosphoramidite, the final DNA templates also contains two constant primer binding sites for PCR amplification, a constant terminator codon to conclude the polymerization, and a tetrathymidine spacer between the coding region and the 3'-end primer binding site. The terminator codon, CGATT, has two G/C bases and three pyrimidines which is identical with all library-comprising codons, is to install a terminator equipped with a fluorescent dye or a biotin group for the characterization or purification of the full-length product after the translation step. The terminator codon was also screened by a MATLAB program to ensure no out-of-frame annealing caused by its presence in the DNA templates.

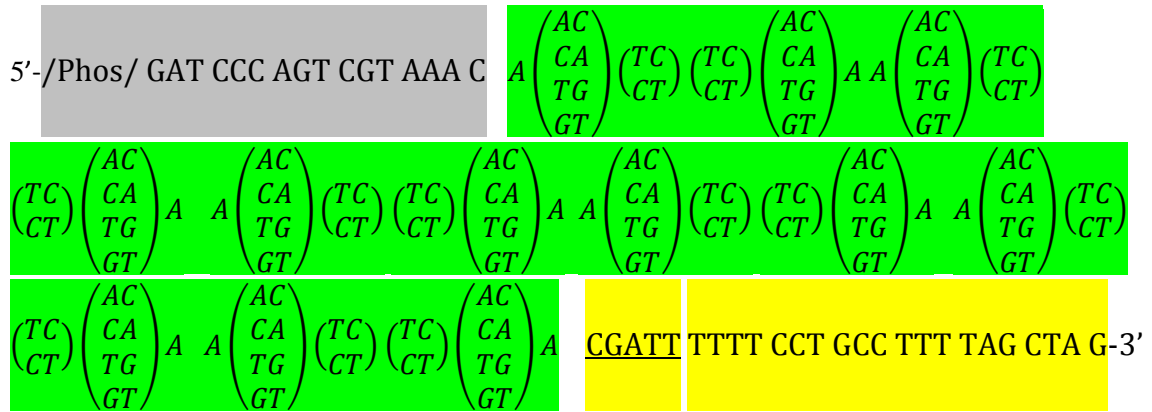


Figure 3.2. Sequence of **LibX**. The grey sequence shows the 5'-end primer binding site and phosphorylation; the yellow sequence shows the 3'-end primer binding site, the terminator codon (underlined), and the tetrathymidine spacer between terminator and the 3'-end primer binding site; the light green region is the variable sequence constructed using the combinations of monomers and dimers. Each codon in the variable region is a five-base sequence, following a pattern of A-dimer-dimer or dimer-dimer-A. The dimers in parentheses are to be incorporated as an equimolar mixture of all dimers within.

Finally, to characterize the quality of the genetic code design and library synthesis, we performed high-throughput sequencing of **LibX** with an Illumina Miseq DNA sequencer. Briefly, the DNA template library was PCR amplified using Illumina sequencing adapters fused with the library primers to install the sequencing adapters. The PCR products were gel-purified, and quantitated using two independent methods: a Picogreen fluorescent assay and a KAPA qPCR assay. The quantified DNA sample were diluted to 4 nM concentration and loaded on Miseq for high-throughput sequencing. The high-throughput DNA sequencing result suggests that designed codon diversity at each codon position of the DNA template coding region is achieved in the library synthesis. Although certain codons, such as ATGCT and TCGTA are more enriched than the others, likely due to slightly higher coupling efficiency of TG and GT dinucleotide phosphoramidites than those of AC and CA dimer, all possible codons have been represented in each corresponding codon position, and the differences between the highest represented codons and the lowest ones are within two fold (Figure 3.3a and b). Moreover, the vast majority of unique DNA templates in the library occurred only once in the two and half million “perfect” sequencing reads, with only 5,800 duplicate reads, highlighting the well-distributed diversity of the library. A perfect read means the sequencing read with the correct read length consistent with the predicted length of the DNA input.

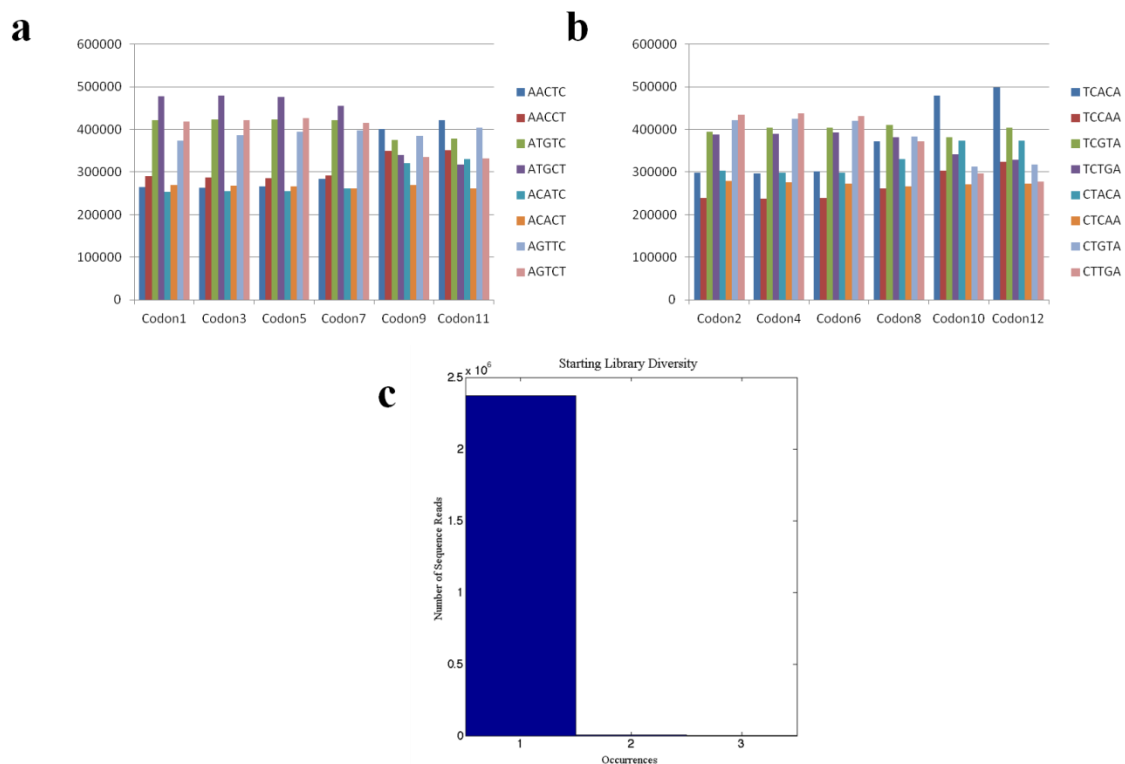


Figure 3.3. (a) Distribution of eight male codons at each of the male codon position. (b) Distribution of eight female codons at each of the female codon position. Codon n , $n = 1, 2, 3, \dots, 12$, indicates the codon positions counting from 5'-end to the 3'-end of the coding region, with codon1 at the 5'-end and codon12 at the 3'-end. (c) Distribution of occurrences of all the perfect sequencing reads. It can be seen that the vast majority of perfect reads are unique sequences.

3.3 Design and Synthesis of the Polymer Building Blocks Comprised of β -Peptides

With the genetic code determined, we start to investigate the synthetic polymer building blocks. In principle, the DNA-templated translation strategy described in Chapter Two is a general method to incorporate synthetic polymer building blocks with various backbone and side-chain structures, so long as they are compatible with the solid-phase chemistry with which the macrocyclic substrates are synthesized, and also compatible with the DNA-templated reaction conditions. However, to be evolved by *in*

in vitro selections in practice the synthetic polymer building blocks are subjected to more requirements. First, the building blocks need to have defined structure and sequence. In order to achieve a fully defined sequence of the translation product, a prerequisite to maintain the heredity of the functions gained by a functional polymer in iterative selection rounds, the synthetic polymer building blocks as reagents must first be sequence-defined. Since most of the synthetic oligomers obtained from chemical polymerization methods such as radical polymerization are not sequence-defined,² they cannot be used as building blocks for translation. In contrast, peptidomimetics, such as β -peptides,³⁻⁶ peptoids,⁷ and other types of synthetic foldamers,^{8,9} are synthetic analogs of natural peptides and mostly synthesized with defined sequence and functional side chains, making them good candidates for synthetic polymer building blocks of DNA-templated translation. Second, the incorporation of synthetic polymer building blocks in the solid-phase synthesis of macrocyclic substrates needs to yield practical amount of substrates for translation reaction. Since some synthetic foldamers were mainly synthesized in solution-phase reactions and may not have protective group schemes compatible with solid-phase synthesis, the incorporation of these building blocks still require further developments before they can be used in our translation strategy. Finally, predispositions of polymer folding for certain types of building blocks make them favorable choices of an *in vitro* selection system. A completely random, unstructured polymer library might require a prohibitively large library size to cover the vast sequence space. As a result, the vast majority of the members in such a library are likely to be very remote from maximum in a fitness landscape; a library of practical size may not even include one

functional member or simply need too much steps to enrich such a member. In contrast, polymers containing structural elements predisposed for chain folding may have a greater chance to emerge in the evolution, because a pre-folded macromolecule has far less entropic cost to form the three-dimensional shape required for functions like binding to a molecular target or catalyzing a chemical transformation.

It is because of these reasons that β -peptides are good candidates for the synthetic polymer building blocks of our translation system. Different from proteinogenic α -amino acids by just one methylene unit, the β -amino acids are still amendable for solid-phase synthesis. But unlike α -peptides, β -peptides cannot be synthesized by ribosomal translation, resulting in virtually no laboratory evolution that has been done with β -peptides. Despite of this non-evolvability, the folding potential of β -peptides has been subjected to substantial amount of study in the last decade.^{3-6,10-12} People now know that β -peptides can fold into secondary structures such as helices and turns, following rules different from α -peptide folding.^{9,13-15}

In Chapter Two we have described a proof-of-principle DNA-templated translation to generate synthetic polymers containing up to 16 hexameric β -peptide building blocks. Based on this result, we further design the β -peptide sequences with various different properties as the synthetic polymer building blocks for the library translation (Table 3.3). A total number of 16 different hexameric β -peptide synthetic polymer building blocks can be further divided into four categories: (1) helix-forming building blocks, including nine sequences following the previously established rules of

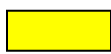



helix formation;^{4,6,15,16} (2) hydrophobic building blocks, including two hydrophobic sequences that may assist hydrophobicity-driven polymer folding; (3) turn-forming building block, including one sequence following the previously established rules of turn formation;¹⁷ (4) diversity-enhancing building blocks, including four sequences that incorporate more different side-chain functional groups than others. All except two β -amino acids involved in building these sequences are β^3 -homoamino acids, as they are the most widely used monomers to form β -peptide secondary structures and also commercially available with the largest collection of functional side chains among all β -amino acids. The exceptions are β^2 -(*S*)-homovaline, a β^2 -monomer that is reported in literature to assist turn formation,¹⁷ and 1*R*,2*S*-aminocyclohexylcarboxylic acid (ACHC), a cyclic monomer that can help induce the helix formation.¹⁶

Table 3.3. β -Peptide building blocks

Azide					Alkyne				
	DNA	PNA	β -peptide	Mw		DNA	PNA	β -peptide	Mw
	codon	anticodon				codon	anticodon		
Lib1	AACTC	GAGTT	flkele	2997.3	Lib2	TCACA	TGTGA	ylkfle	3199.6
Lib3	AACCT	AGGTT	eavdki	2893.2	Lib4	TCCAA	TTGGA	elsklq	3104.5
Lib5	ATGTC	GACAT	fewelk	3039.4	Lib6	TCGTA	TACGA	fklyel	3168.6
Lib7	ATGCT	AGCAT	rgydle	2807.1	Lib8	TCTGA	TCAGA	el(ACHC)kle	3098.5
Lib9	ACATC	GATGT	wsv'yew	3088.4	Lib10	CTACA	TGTAG	nlktle	3104.5
Lib11	ACACT	AGTGT	lhdlyl	3063.4	Lib12	CTCAA	TTGAG	qkliel	3130.6
Lib13	AGTTC	GAAC	elkhle	3027.4	Lib14	CTGTA	TACAG	khsegq	3004.4
Lib15	AGTCT	AGACT	lkdleq	2982.3	Lib16	CTTGA	TCAAG	hywkfe	3336.7

Table 3.3 (Continued)

*All β^3 -amino acids are represented using the corresponding α -amino acid one-letter abbreviation with the same side chain, but with lower-case letters. v': β^2 -(*S*)-homovaline. ACHC: 1*R*,2*S*-aminocyclohexylcarboxylic acid

	Helical elements
	Hydrophobic elements
	Diversity-enhancing elements
	Turn element

All substrates containing the β -peptide building blocks were synthesized following the procedure described in Chapter Two. The β -peptides were coupled using the Fmoc N-protected β -amino acid monomers with side-chain protections same as the corresponding α -monomers commonly used in solid-phase peptide synthesis. Since the coupling reactions of β^2 -(*S*)-homovaline and ACHC have been proven to be difficult,^{17,18} double couple with extended two hour coupling time were used for these monomers and the monomers coupled immediately after them. All substrates have been extensively purified by HPLC until they show a single peak consistent with their expected masses on a MALDI spectrum.

3.4 Re-Optimization of the Translation Reaction Condition to Facilitate Library Translation

The translation reaction conditions described in Chapter Two have been successfully used to perform translation of a single DNA template with two alternately

repeating codons. Unfortunately, the same reaction conditions are not successful when applied to the translation with mixed codons (Figure 3.4a, first gel), as no full-length product can be made under the previous condition. This discrepancy is likely caused by more complex mixture of templates and substrates in the translation reaction, , in which a DNA template library containing numerous templates with mixed codons and all 16 substrates are mixed. The non-specific interactions among substrates may cause undesired complexation or even aggregation, lowering the effective concentration of these substrates to participate in the templated reaction or causing early truncations during a polymerization process. In addition, the new reaction condition also poses a challenge for the purification of the translation products, as the increased amount of non-reacted substrates and copper-ligand complex tend to evade the size-exclusion purification, causing inefficient adapter cleavage and reduced overall stability of the translation polymers. Therefore, a purification method with higher efficiency in removing substrates and catalysts that are not attached to the DNA templates is greatly needed.

To overcome these problems, we sought to re-optimize the translation reaction condition first for translation using a single template with mixed codons. A 12-codon DNA template **Dm1**, containing codon AACCT, TCCAA, TCGTA, ATGCT, ACATC, CTACA, ACACT, AGTCT, CTTGA (some codons appear more than once), and terminator codon TACCA, was used to test the reaction condition in the presence of all 16 substrates **Lib1~Lib16**, and terminator **T17f** (see Method section for terminator structures). To reduce the non-specific interaction among substrates, we tested anionic, cationic, and non-ionic detergents in the translation reaction mixture. Not unexpectedly,

anionic and cationic detergents interfered with the DNA template-substrate hybridization and resulted unsuccessful translation reaction, probably due to their ability to bind to the DNA backbone or nucleobases. Fortunately non-ionic detergents such as Tween-20 and 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS, Figure 3.4b) were able to reduce the non-specific interaction among substrates and improve the yield of full-length products (Figure 3.4a, second gel). To simplify the purification, we chose to use CHAPS for further study since it forms smaller micelles than other non-ionic detergents, resulting in easier removal by size-exclusions.

Next, we studied the stoichiometry between the substrates and the DNA template. If non-specific interaction could decrease effective concentration of the substrates, increasing the overall concentration of the substrates in the reaction should offset this decrease. Using **Dm1**, terminator **T17f**, and all 16 substrates **Lib1~Lib16**, we tested three different substrate/template ratios, adding five, ten, or 15 equivalents of substrates per equivalent of DNA template. The translation of **Dm1** in the presence of 15 equivalents of substrates gives higher yield of the full-length product compared to the reaction with five equivalents of substrates (Figure 3.4a, the third gel from left), suggesting a successful translation of a DNA template of mixed codons require higher dosage of substrates to remedy the loss due to non-specific substrate interactions.

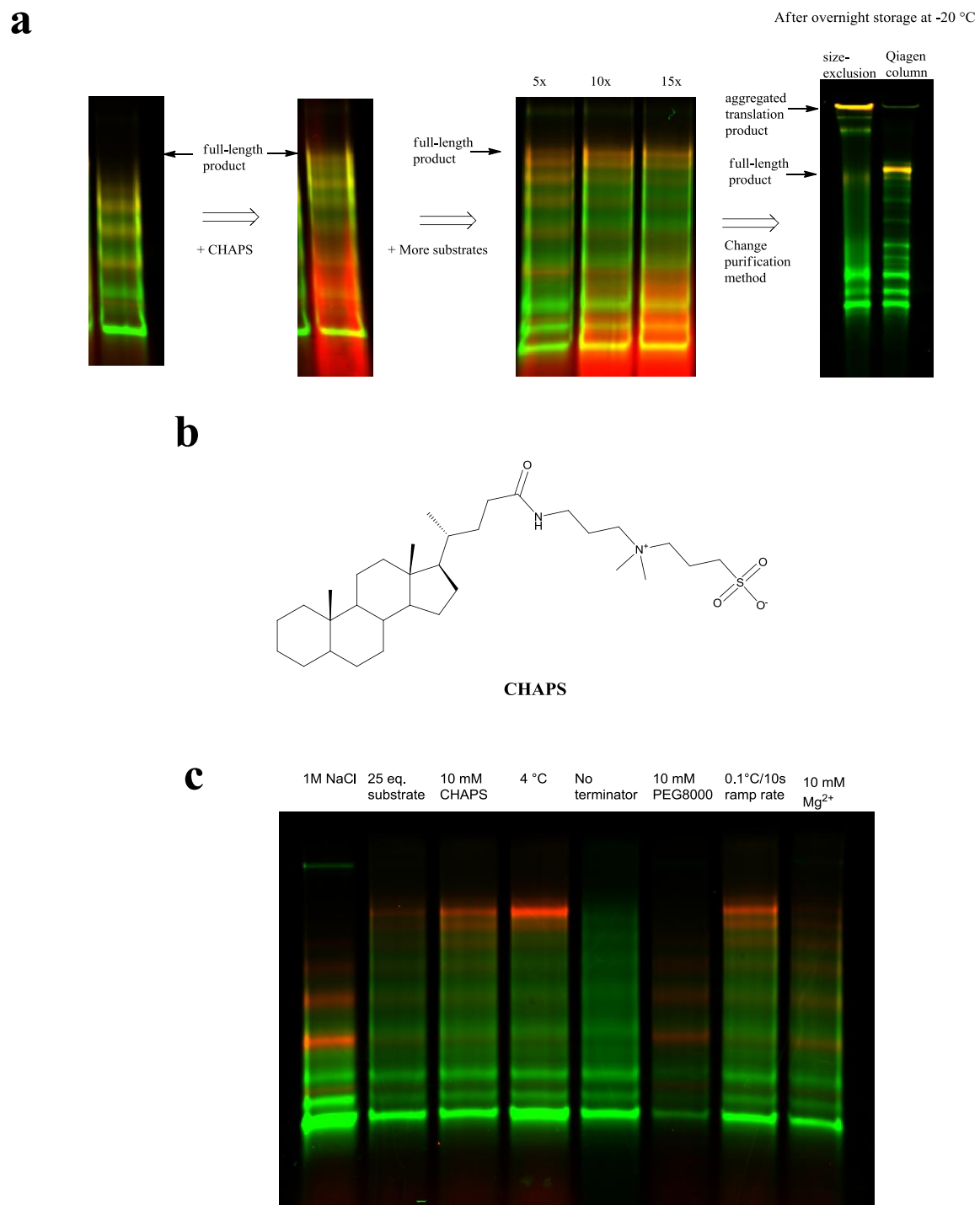


Figure 3.4

Figure 3.4 (Continued)

Figure 3.4. Re-optimization of the translation reaction condition for library translation. (a) Optimization reaction conditions for the translation of a single DNA template containing mixed codons. DNA template **Dm1**, terminator **T17f** (see Method section for structure of terminators) and all 16 substrates shown in Table 3.3 were used in translation. Starting from a reaction condition that does not yield full-length polymerization product, the translation efficiency was improved by adding CHAPS to reduce non-specific interaction among substrates. The yield of the full-length product was further improved when 15 eq. of substrates to the DNA template were used, generating the full-length product as a major product. The side-reactions during adapter cleavage and storage were further reduced when the size-exclusion purification method was replaced by the Qiagen spin column (Minelute kit). The translation product after Qiagen column purification shows a cleaner result on PAGE analysis, and became more stable over long-term storage. (b) Structure of the detergent CHAPS. (c) Optimization of the library translation. The DNA template library **LibX**, all 16 substrates from Table 3.3, and terminator **T18f** were used in translation. We tested eight variations on the basis the reaction conditions obtained from single mixed-codon template translation, including additional 1 M NaCl salt, 25 eq. of substrates to the DNA template, 10 mM CHAPS, 4 °C incubation overnight, no terminator added, 10 mM PEG8000 molecular crowding agent, 0.1 °C slow ramp rate during annealing step, and 10 mM divalent Mg^{2+} ion. All the gels were imaged by a Typhoon fluorescent imager with two channels being imaged. The green channel shows Cy3 dye with which the DNA template is labeled. The red color shows Alexa Fluor 647 with which the polymerization terminator was labeled.

Finally, as more substrates added in the translation reaction, the removal of catalysts, ligands and un-reacted substrates by the size-exclusion purification becomes more challenging. A significant amount of un-reacted substrates leaked through size-exclusion purification will consume the reductant of the subsequent disulfide cleavage step, causing incomplete disulfide reduction and/or incomplete thiol alkylation. Moreover, the copper catalysts and the ligands evading the purification may also cause problems,

because copper(I) ion and its oxidized version, copper(II) ion could catalyze DNA damage and polymer backbone and side-chain oxidation. To avoid these problems, we attempted to replace size-exclusion columns with Qiagen spin columns to purify the translation reaction. Unlike the size-exclusion columns such as Princeton Separations, the one we have been using for translation purification, Qiagen spin columns use a proprietary silica membrane to selectively absorb DNA in the presence of high concentration of chaotropic salts under low pH. The bound DNA can be thoroughly washed to remove impurities before elution under neutral pH. This strategy is likely to ensure removal of all the contaminants in the translation mixture, leaving only translated products covalently linked with the DNA templates after purification. As expected, the translation products purified by Qiagen spin columns were much cleaner, with no non-templated by-products that were inefficiently removed by size-exclusion methods were seen on the gel. After disulfide cleavage and thiol alkylation, the band corresponding to the full-length product was also sharper than previously (Figure 3.4a, the fourth gel from left). Importantly, the translation products that have been purified by Qiagen spin columns are more stable and can be safely stored under -20 °C. In contrast, the reactions purified by previous size-exclusion methods always require an immediate PAGE purification after adapter cleavage, otherwise the products tend to form high molecular weight aggregates over storage that cannot even migrate into the gel during PAGE analysis.

Lastly, we optimized the reaction condition for library translation. The translation using DNA template library **LibX**, terminator **T18f**, and all 16 substrates

Lib1~Lib16 were further optimized on the basis of the reaction condition determined in the translation of a single mixed-codon template. Eight different variation of reaction conditions were introduced, including additional 1 M NaCl salt (the original condition has 0.2 M NaCl), 25 eq. of substrates to the DNA template, 10 mM CHAPS, 4 °C incubation overnight, no terminator, 10 mM PEG8000 molecular crowding agent, 0.1 °C/10s slow ramp rate during annealing step, and 10 mM divalent Mg^{2+} ion (Figure 3.4c). The PAGE analysis of translation results suggests three conditions, namely 4 °C reaction temperature, 10 mM CHAPS, and 0.1 °C/10s slow ramp rate during annealing step, can further improve the yield of full-length products, as the fluorescent intensities of the correctly terminated polymer chain increase under these reaction conditions. Therefore, we added these three conditions to the reaction conditions optimized for a single mixed-codon template to be the finalized reaction condition for library translation.

3.5 Study of Sequence-Specificity of Library Translation

With the genetic code design determined, all 16 substrates **Lib1~Lib16** synthesized, and the reaction condition re-optimized for library translation, we wonder if the sequence-specificity still remain unchanged from the model system of Chapter Two. We designed two DNA templates **Ds1** and **Ds2** to test the sequence specificity of the library translation. **Ds1** contains codons AACTC, TCACA, AACCT, TCCAA, ATGTC, TCGTA, ATGCT, TCTGA for the DNA-templated polymerization of **Lib1~Lib8**; **Ds2** contains codons ACATC, CTACA, ACACT, CTCAA, AGTTC, CTGTA, AGTCT, CTTGA for the DNA-templated polymerization of **Lib9~Lib16**. We also designed and

synthesized 16 terminators **TLib1~TLib16** (Table 3.4), each is able to terminate the polymerization at a specific codon position of **Ds1** and **Ds2**. To test the sequence-specificity of the translation, we performed the specific termination experiment similarly as the one described in Chapter Two, Section 2.5. We attempted 16 translation reactions, each of which contains **Ds1** or **Ds2** template, a terminator from **TLib1~TLib16**, and seven substrates corresponding to the non-terminating codons from **Lib1~Lib16**. Only if the terminators and the other substrate are all installed sequence-specifically, opposite to their complementary codons, will the polymerization generate a product of the desired length.

Table 3.4. Terminators for Sequence-Specificity Test

Azide					Alkyne				
	DNA	PNA	Polymer	Fluorescent		DNA	PNA	Polymer	Fluorescent
	codon	anticodon	building	Dye		codon	anticodon	building	Dye
			block					block	
TLib1	AACTC	GAGTT	2× PEG ₂	AF647	TLib2	TCACA	TGTGA	2× PEG ₂	AF647
TLib3	AACCT	AGGTT	2× PEG ₂	AF647	TLib4	TCCAA	TTGGA	2× PEG ₂	AF647
TLib5	ATGTC	GACAT	2× PEG ₂	AF647	TLib6	TCGTA	TACGA	2× PEG ₂	AF647
TLib7	ATGCT	AGCAT	2× PEG ₂	AF647	TLib8	TCTGA	TCAGA	2× PEG ₂	AF647
TLib9	ACATC	GATGT	2× PEG ₂	AF647	TLib10	CTACA	TGTAG	2× PEG ₂	AF647
TLib11	ACACT	AGTGT	2× PEG ₂	AF647	TLib12	CTCAA	TTGAG	2× PEG ₂	AF647
TLib13	AGTTC	GAAC T	2× PEG ₂	AF647	TLib14	CTGTA	TACAG	2× PEG ₂	AF647
TLib15	AGTCT	AGACT	2× PEG ₄	AF647	TLib16	CTTGA	TCAAG	2× PEG ₄	AF647

For all 16 reactions, the major product was the polymer of the desired length, as indicated by the fluorescent intensity of the band with overlapping fluorophores (Figure

3.5). This result confirms that the recognitions of the DNA template codons and the substrate PNA anti-codons designed as the genetic code are sequence-specific, reassuring the fidelity of the library translation.

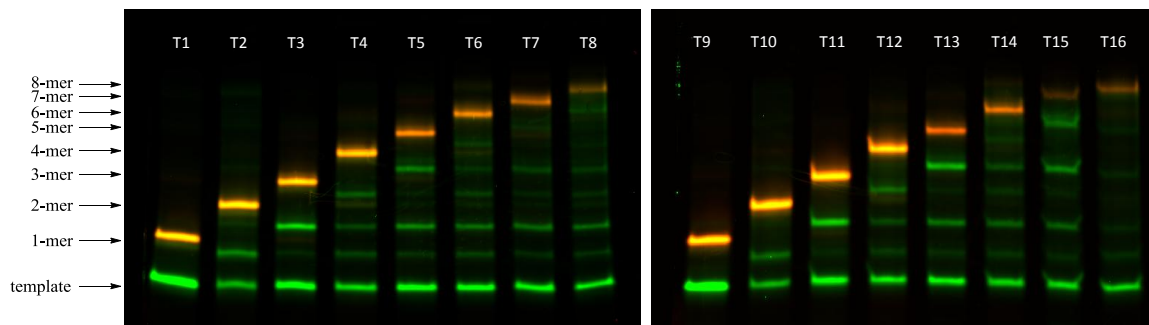


Figure 3.5. Each reaction uses a Cy3 (green channel) labeled DNA template, **Ds1** or **Ds2**, each containing eight codons from the PNA genetic code. Each translation contains a terminator labeled with Alexa Fluor 647 (red channel) and seven bifunctional substrates from **Lib1~Lib16**. The incorporation of the terminator stops the chain growth and generates a polymer product indicated by the overlapping band of the two fluorescent channels. The polymer lengths reflect the sequence-specific incorporation of terminators, thereby demonstrating the sequence-specificity of the translation. Tn , $n = 1, 2, \dots, 16$ indicates the terminator **TLib1~TLib16** being used in the reaction.

3.6 Study of Translation Using DNA Templates with Various Propensities of Secondary Structures

In order for the sequence-defined synthetic polymer library to minimize translation bias, ideally all members of the DNA template library should be translated to its corresponding synthetic polymer translation product. However, as we have discussed above, potential DNA folding secondary structures in the coding region of a DNA template could render the single-stranded coding region inaccessible for the substrates,

thereby inhibiting the translation reaction using that template. If large amount of such secondary structures of coding region exist in the DNA template library, translation bias will emerge in favor of those sequences with low secondary structure propensity, while biasing against those with high secondary structure propensity. In Section 3.2, we have integrated the criterion of low secondary structure propensity in the search of synthetic polymer genetic code, and obtained a codon set with relative low secondary structure propensity. More than 95% of the DNA templates constructed using the DNA codons in this genetic code were predicted to have secondary-structure energies below -5.0 Cal/mol under the ambient folding conditions, rendering them only weakly structured. To test if the DNA templates in this genetic code design can indeed be translated successfully, we sampled the DNA template library with templates of different secondary-structure energies, ranging from very high and very low propensities of DNA secondary structures, and performed translation using these templates.

Pleasantly, we found all translations using DNA templates with various secondary-structure energies proceeded successfully to generate full-length products (Figure 3.6). The template with the highest secondary structure folding propensity, **Df2**, has a secondary-structure energy of -7.01 kCal/mol, higher than 99.9% of the templates in the library. Yet the translation using **Df2** successfully generated full-length translation product. Notably, the translation efficiencies are not always inversely proportional to the secondary-structure energies. **Df15**, a DNA template having no secondary structure predicted, generate the full-length product in a lower yield in translation compared to DNA templates with higher secondary-structure energies. Such a result suggests while

the coding region secondary structures of the DNA template library may not pose significant challenges in translation, many factors other than the coding region secondary structures could influence the translation efficiency.

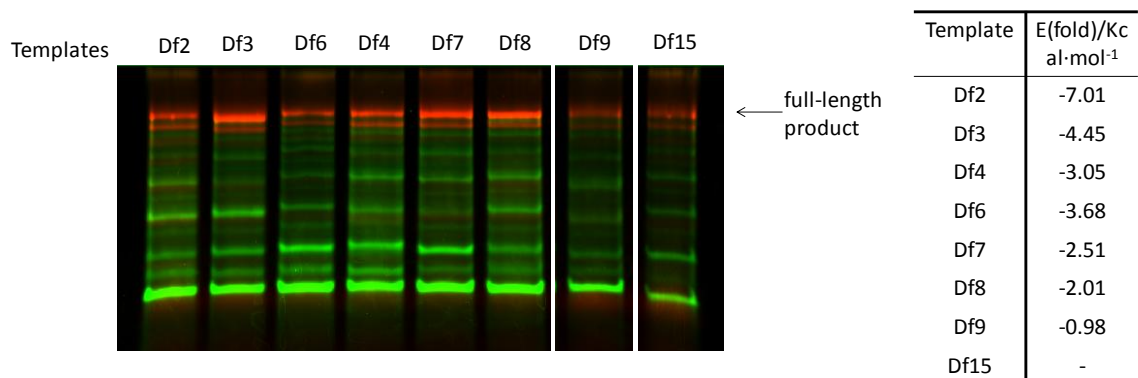


Figure 3.6. Translation of DNA templates with various secondary-structure energies. Each translation reaction contains the terminator **T18f**, which is installed at the end of the translation products by the terminator codon and is labeled with Alexa Fluor 647 (red channel). The secondary-structure energies of the templates used are shown in the right table.

3.6 Developments towards A Primer Extension Strategy and An Affinity

Purification Strategy of the Full-Length Product

To avoid the DNA components of the translation product fold into conformations that help the survival of this library member independent of the synthetic polymer component, we reason that by making the DNA template after translation double-stranded could mask its potential ability of folding and functioning other than an information carrier. We chose Klenow fragment (3'→5' exo-) to catalyze the primer extension of the 3'-end primer of the DNA template, because this DNA polymerase, unlike many other polymerases for PCR reaction, functions at a milder temperature of

37 °C and does not need thermal cycling. This property could help reduce the disruption of the DNA template library and the risk of side-reactions on polymers at high temperature. Furthermore, the Klenow fragment (3'->5' exo-) has strong strand displacement capability, which is potentially useful in the displacement of the template hairpin region to synthesize the complementary strand that covers the whole template.

Following translation using a single mixed-codon DNA template **Dm1**, terminator **T17f**, and substrates **Lib1~Lib16**, the translation product purified by a Qiagen spin column (Minelute kit) was added the reverse primer (binding to the 3'-end primer binding site) and Klenow reaction buffer. The reverse primer was extended at 37 °C by Klenow fragment (exo-) in the presence of dNTPs (Figure 3.7a). Another Qiagen spin column purification (Minelute kit) was required to remove the enzyme and other DNA polymerization ingredients, as Klenow fragment (3'->5' exo-) kept associated with the translation product if there was no further cleanup after the primer extension reaction. After successful primer extension and enzyme removal the product returned a ladder pattern on the native PAGE, with the topmost band having overlapping fluorescent dyes to indicate the full-length translated polymer. It should be noted that due to the unique nature of the translation product as a DNA-synthetic polymer conjugate, after primer extension its mobility on a native PAGE increases, moving faster than before the reaction (Figure 3.7a). The same phenomenon was also observed previously in an mRNA-display translation system, in which the translation product was also a double-stranded DNA-polypeptide conjugate.¹⁹

Next, we investigated the methods to purify the full-length translation product. Since the library translation leaves a significant amount of the non-translated template and pre-mature truncations, it is desirable to remove these by-products before entering the translation products in the selection step of an *in vitro* selection cycle. Removal the imperfect translation products can also help simplify the interpretation of the *in vitro* selection results, ensuring that one DNA template only encodes one full-length, sequence-specific translation product.

We reason that by installing a cleavable biotin group on the terminator that is incorporated at the end of the translated polymer, enrichment by streptavidin immobilized affinity medium of only the full-length translation product can be achieved. Since a constant terminator codon CGATT has been placed at the end of the coding region of **LibX** during the chemical synthesis of the library, in principle all the translation product of the library is able to be purified independent of the sequence context of the library members using the affinity purification strategy.

We first attempted the affinity purification of the full-length product from a translation using a single mixed-codon DNA template **Dm1**, substrates **Lib1~Lib16**, and a terminator **T17b** that is equipped with a biotin group bridged by a disulfide cleavable linker (see Method section for detailed structure of **T17b**). Immediately after DNA-templated polymerization and Qiagen spin column purification, and prior to adapter cleavage, the polymerized material was subjected to streptavidin immobilized purification medium. By using a disulfide linker to link the terminator and the biotin

group, the adapter cleavage of the substrate and the elution of the streptavidin bound full-length translation product can be effected by a single treatment of disulfide cleaving reductant such as DTT or (tris(2-carboxyethyl)phosphine) (TCEP) and subsequent thiol alkylation. As expected, the affinity pull-down by streptavidin magnetic particles and the subsequent reductive release of the full-length translation products are effective, with the full-length products significantly enriched (Figure 3.7b, lane four). Further analysis of the flow-through of the streptavidin affinity purification indicated that only a small fraction of the full-length product escaped the affinity pull-down (Figure 3.7b, lane three), highlighting the efficiency of the affinity purification strategy.

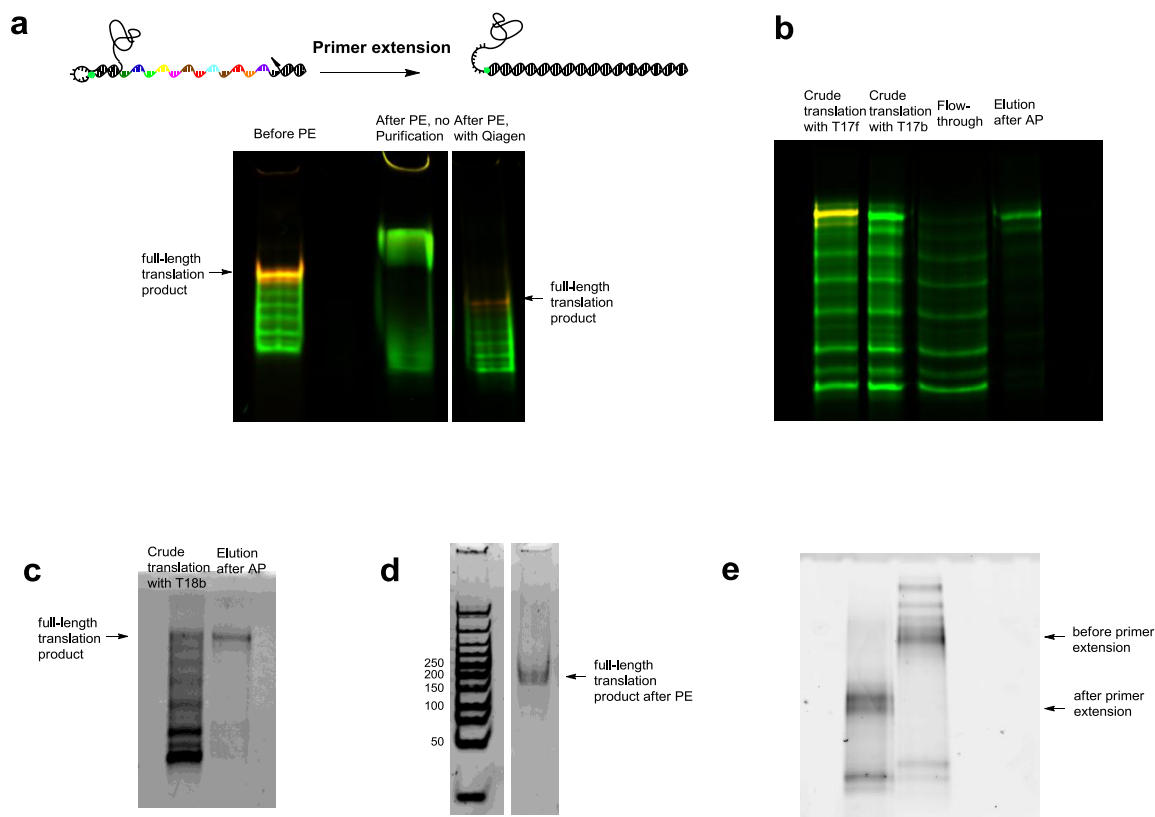


Figure 3.7. Developments of primer extension and affinity purification strategies. (a) Primer extension of a translation product from reactions using DNA template **Dm1**, substrates **Lib1~Lib16**, and terminator **T17f**. (2) Affinity purification is effected by streptavidin magnetic particle capture of the full-length translation product from the reaction using DNA template **Dm1**, substrates **Lib1~Lib16**, and terminator **T17b**. (c) Affinity purification of the library translation using DNA template **Dm1**, substrates **Lib1~Lib16**, and terminator **T18b**. (d) Subsequent primer extension of the translation product purified in c. (e) A SDS-PAGE analysis of the library translation products before and after primer extension. PE: primer extension; AP: affinity purification. **T17f** is covalently attached with an Alexa Fluor 647 fluorophore, **T17b** and **T18b** both have a biotin side chain bridged by a disulfide linker. **T17** and **T18** terminators can be installed at the last codon position of **Dm1** and **LibX**, respectively. The structures of **T17f**, **T17b**, and **T18b** can be found in the Method section.

We further tested the combined efficiency of the affinity purification step and a subsequent primer extension step. Translation using the DNA template library **LibX**, substrates **Lib1~Lib16**, and a terminator **T18b** was purified by a Qiagen spin column (Minelute kit), and then subjected to streptavidin magnetic particles for affinity purification. After disulfide cleavage and full-length translation product elution, the eluted material was analyzed by denaturing PAGE. Although the full-length yield of the library translation is lower than that of a single DNA template translation, the affinity purification method was still able to significantly enrich the full-length translation product and remove most of the imperfect by-products (Figure 3.7c). The purified material was also subjected to a primer extension reaction catalyzed by Klenow fragment (exo-). A single, slightly expanded band of the expected size corresponding to primer-extended product on a native gel indicates the combined approach of affinity purification and the primer extension yielded the full-length translation product with the DNA component double-stranded (Figure 3.7d). As the native gel resolve poorly the translation product before primer extension, possibly due to the single-stranded DNA template region and the presence of the synthetic polymer, we also performed SDS-PAGE analysis of the library translation product before and after primer extension. The SDS detergents bind to the synthetic polymer region of the translation product and denature them, resulting more homogeneous migration in gel. SDS-PAGE reveals that the primer extension was complete, and that after primer extension the translation products migrate faster than before primer extension, which is consistent with the single-template primer extension result analyzed by the native PAGE (Figure 3.7e).

We evaluated quantitatively the final yield of this desired product by qPCR. Quantitative qPCR results suggest that the yield of affinity purification is 2% of the input DNA template library, while after primer extension the yield of the final product drops to 0.9%. The low yield of the full-length final product is likely due to the combined factors of lowered translation efficiency of a library translation, material loss in the affinity purification step, and the loss during two Qiagen spin column purifications that are required to prepare the materials for affinity purification and primer extension. Although likely to cause higher consumption of the DNA template library and substrates, a potential solution to this problem is to scale up the translation and ensure the resulting material has enough molecules to achieve a good coverage of the entire library.

3.7 A Mock Selection: Iterative Rounds of Translation, Enrichment of Biotinylated Polymer by Streptavidin Capture, Template Re-Generation, and Re-Translation.

Iterative cycles of *in vitro* selection and amplification can allow extremely rare but highly functional species to emerge from a structurally and sequence-diverse library.²⁰⁻²² In order for the sequence-defined synthetic polymers to undergo iterative *in vitro* selection cycles, the DNA template encoding these polymers must be amplified and re-translated. We sought to first test the ability of the enzyme-free translation system to support PCR amplification using the translation product as the template, template re-generation, and re-translation. We designed a DNA template **D11** that can be amplified using two modified primers (Figure 3.8a). One of the primer contains a 5' hairpin loop that installs a 5' hexynyl group for coupling with the synthetic polymer; the other primer

is equipped with a 5' biotin group to be used in a downstream strand separation. After translation using substrates **22**, **23**, and terminator **T7** (see Chapter Two, Section 2.8, Table 2.1 and 2.3 for structures of these substrates), the full-length DNA-synthetic polymer conjugate was purified by PAGE (Figure 3.8a, lanes 3 and 4). 1/50,000th of the resulting material was subjected to PCR amplification (Figure 3.8a, lane 5). Sanger sequencing of the PCR product revealed that the sequence of the starting DNA template was identical after translation (Figure 3.8b). The double stranded PCR product underwent strand separation and the sense strand was eluted after alkaline treatment of the streptavidin immobilized DNA (Figure 3.8a, lane 6). The resulting single stranded DNA was able to serve as a template for re-translation using the same substrates to generate translation products that are indistinguishable by PAGE electrophoresis from the products generated by the starting template (Figure 3.8a, lane 7). Taken together, these results validate that the DNA-templated translation strategy is able to be integrated in a complete cycle of translation, amplification, and template re-generation.

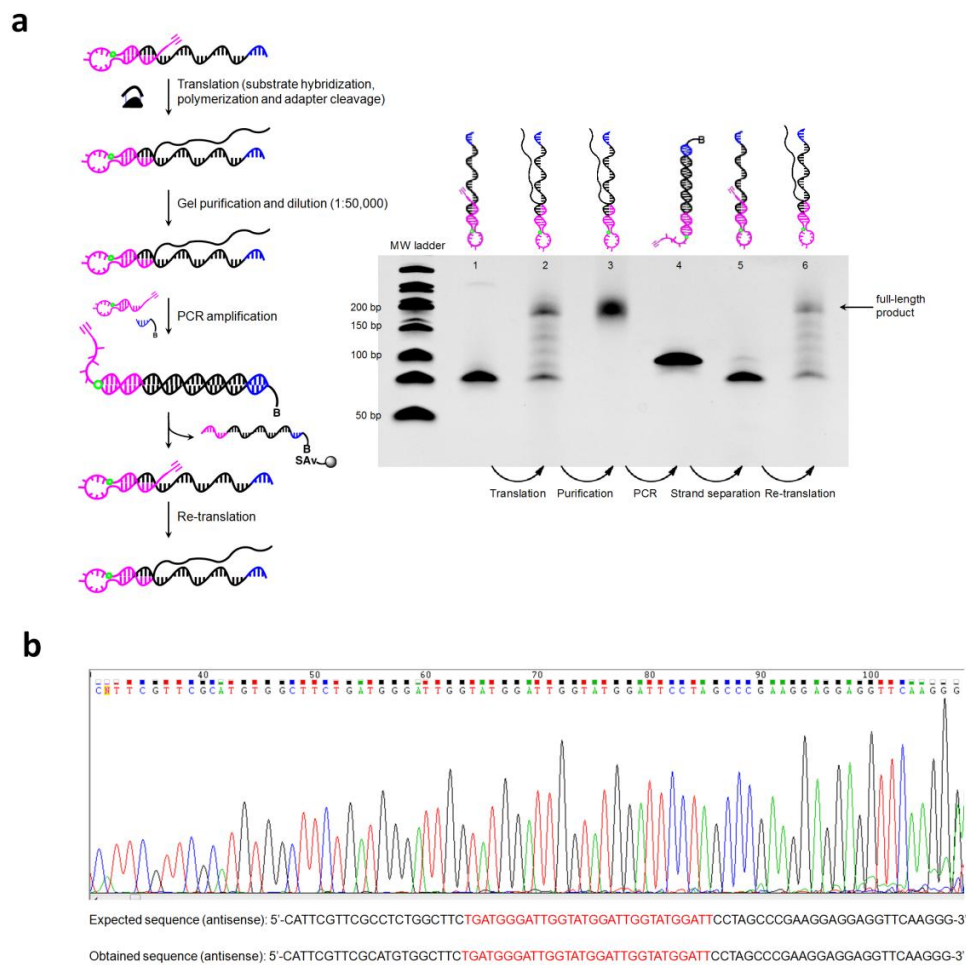


Figure 3.8. (a) A 100 pmol starting template **D11** was translated into synthetic β -peptide polymer using substrates **22**, **23** and **T7** (lanes 1 and 2). The translation reaction was purified by denaturing PAGE (lane 3). A small amount (1/50,000th) of the translation product, simulating the amount surviving a typical in vitro selection, was subjected to PCR with two primers, generating a double-stranded template (lane 4). The PCR product was immobilized on streptavidin-linked beads and the ‘sense’ strand was eluted with 150 mM NaOH (lane 5). Finally, this single-stranded template was re-translated into synthetic polymer using the same substrates (lane 6). The non-denaturing PAGE gel shown was visualized by ethidium bromide staining. B, biotin; SA, streptavidin. (b) DNA sequence of template DNA amplified by PCR from a minute fraction (1/50,000th) of translation product. DNA sequencing was performed for the antisense strand using a sequencing primer **SeqFP** (5’-TAATACGACTCACTATAGGGCTCGATTTAATTCGCCGACGTGATGACATTCCAGGCAGTCATTCGTTTCGCTCTGGCTTC-3’). The red bases indicate the coding region. DNA sequencing confirms that the translation and amplification processes did not alter the sequence of the DNA template.

Having confirmed that the translation product can be PCR amplified and the PCR product can be used to re-generate the DNA template of the next round of translation, we wonder if an *in vitro* selection system, starting from a DNA template library and including library translation, selection, PCR amplification, template re-generation, and re-translation, could be used to enrich an extremely small fraction of DNA molecules that encode a known function. Importantly, such a mock selection experiment would form the foundation of *in vitro* selection of sequence-defined synthetic polymers with bioactive functions such as binding to a molecular target or catalyzing a chemical reaction.

Because we have already tested the robustness and feasibility of the biotin-streptavidin interaction in the affinity purification of the full-length translation products, to simplify the work flow and avoid the problems brought by interactions not previously verified, we decide to choose affinity enrichment of the biotinylated library members by streptavidin magnetic particles as the selection criterion for the mock selection experiment. A positive control DNA template, **Dpos**, is designed to have the same primer binding sites as the DNA template library **LibX**, and contain a 60-mer coding region of 12 pentameric codons, followed by a biotinylation codon GCATT that in theory should not appear in **LibX** and can encodes the incorporation of a biotinylated terminator (Figure 3.9a). The last three bases of the biotinylation codon and a following three-base spacer, AAT, is designed to form an AseI restriction enzyme cut site, ATTAAT, to allow

the digestion of **Dpos** to be observable when it is sufficiently enriched. We mixed **LibX** and **Dpos** in a 10^6 :1 ratio, so that there has to be significant enrichment of **Dpos** by in vitro selection before its restriction digest can be observed.

Starting from the DNA template library of **LibX:Dpos** = 10^6 :1, a complete mock selection cycle includes the following steps (Figure 3.9b). First, the DNA template library are annealed with substrates **Lib1~Lib16**, and the DNA-templated polymerization undergoes in the presence of Cu(I) catalyst and THPTA ligands. The DNA-templated polymerization products of all LibX library members are designed to be terminated by **T18a**, a terminator with an acetylated lysine side chain (see Method section for the structure of **T18a**). The DNA-templated polymerization product of **Dpos** is designed to be terminated by **T19b**, a terminator with side chain containing a biotin group, bridged by a disulfide linker (see Method section for the structure of **T19b**). Second, after the DNA-templated polymerization reaction the reaction is purified by a Qiagen spin column (Minelute kit). The purified material is subjected to streptavidin magnetic particles to capture the biotinylated products of the polymerization reaction. Third, the captured molecules are eluted after sufficient wash of the magnetic particles, by the treatment of the particles with TCEP containing elution buffer. The eluted material serves as the template in a subsequent PCR to amplify the DNA templates enriched by the selection step, using a 5'-phosphorylated forward primer and a 5'-biotinylated reverse primer. Fourth, the PCR products are strand-separated by binding to the streptavidin magnetic particles and then eluting with alkaline treatment. Note this capture and elution step of the PCR products are for the separation of the two strands of the PCR products; this step

should not be confused with the selection step. Finally, the strand-separated product is ligated with the hairpin region of the template to form the whole template for the next round of translation.

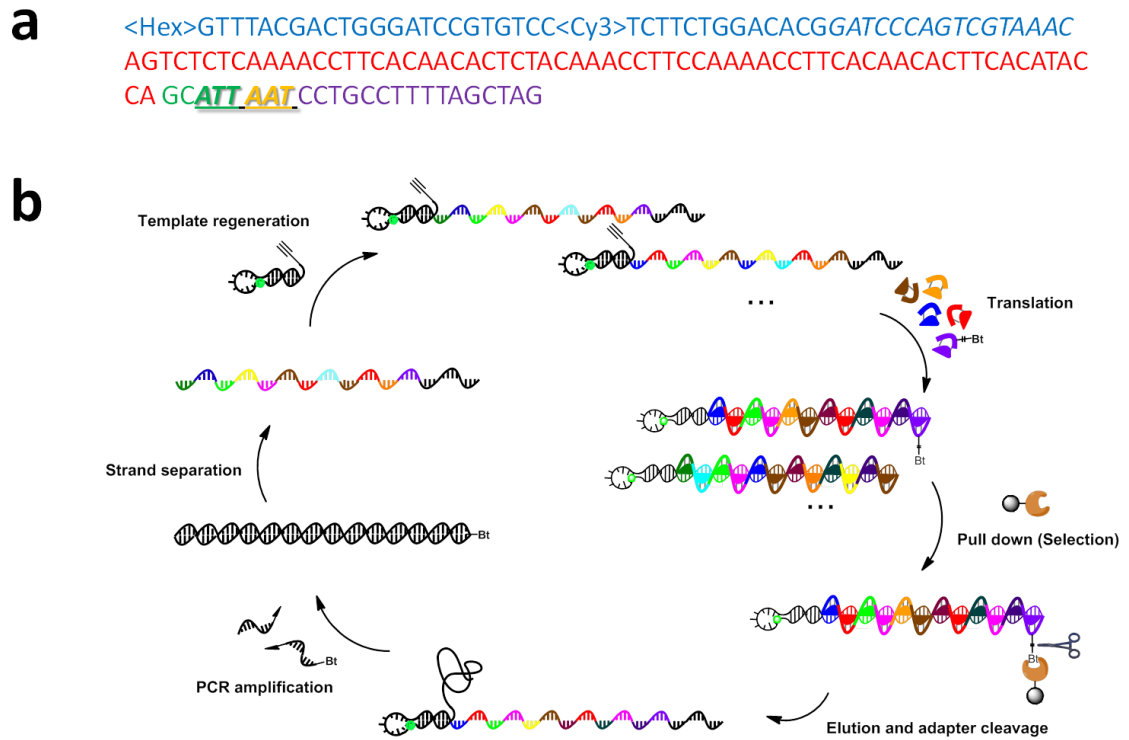


Figure 3.9. (a) Sequence of **Dpos**. Blue sequence indicates the hairpin region; red sequence indicates the coding region; green sequence indicates the biotinylation codon; yellow sequence indicates the spacer; purple sequence indicates the reverse primer binding site. The blue sequence in *Italic* fonts indicates the forward primer. <Hex> and <Cy3> are hexylnyl and Cy3 modifiers. The sequence in bold and underlined indicates the *AseI* cut site. (b) Schematic illustration of the mock selection cycle. Bt: biotin.

We performed six rounds of mock selection to enrich the DNA template library members that encodes biotinylation. At the end of the sixth round, over 12% of the total

products translated from the library were bound to streptavidin magnetic particles, suggesting an enrichment of the biotinylated species (Figure 3.10a). However, the AseI digestion did not show any significant amount of the total DNA library was cut even at the Round Six (Figure 3.10b). We speculate that the reason for these seemingly contradictory and confusing results is that before mixing with **Dpos**, the DNA template library **LibX** may have already had the biotinylation codon GCATT, causing unexpected incorporation of the biotinylated terminator **T19b**. If the number of these “false positive” templates containing unexpected biotinylation templates far exceeds the number of **Dpos** that is mixed with **LibX** in a 1:10⁶ ratio, these DNA templates will reach the exponential growth stage earlier than **Dpos** templates in the library, thereby dominating the enrichment process. As a consequence of the existence of these false-positive templates in **LibX**, the enriched library can generate translation products with significant binding to streptavidin, but does not have enough numbers of AseI cut site and therefore does not show significant digestion.

To test this hypothesis, we first designed two PCR primers **Rpos** (5'-AGGCAGGATTAATGC-3') and **Rfpo** (5'-AGGCAGGAAAAAATGC-3') to probe the sequence of the DNA template library, respectively targeting the AseI cut site of **Dpos**, GCATTAAT (“positive codon”), and a hypothesized mutation of LibX, CGATTTTTT → GCATTTTTT (“false positive codon”), that may have caused the unexpected biotinylation. Cross PCR amplifications of these two primers confirm that they are specific to their designated target sequences, exhibiting 4,000 and 10⁷ fold lower cross-amplifications on the mismatched templates than the amplification on the matched ones.

Using these primers as probes in the qPCR experiments, we quantitatively determined the amount of enrichment of each sequence targeted by the probes. Starting from the DNA template library of Round One, the false positive sequences are at least 30 fold more than the positive codons (Figure 3.10c). As the selection progressed, the false positive codons reached the exponential growth stage earlier than the positive codons, leading to its dominance in the selection (Figure 3.10c). After six rounds, the false-positive codons have taken a significant fraction of the entire library, as opposed to the modest enrichment of the positive codons that was still 1,000 fold less enriched.

The qPCR experiments demonstrate the quantitative difference of the enrichments of **Dpos** and a hypothesized mutation in **LibX**. In order to unambiguously determine if some mutations of LibX that can cause unexpected biotinylation in the translation product indeed dominated the DNA template library, we performed high-throughput sequencing of the DNA template library after Round Six, and compare the result with that of the starting library (Figure 3.10d). High-throughput sequencing results demonstrate that among all sequencing read of the Round Six library, 23% of the reads contain GCATT biotinylation codon. Almost entirely (988,326 out of 990,141), these reads are unexpected biotinylation codons derived from mutations in **LibX**. Only a small fraction (1,805 out of 990,141) of the biotinylation codons are within **Dpos** templates. In addition to the mutation at end of the coding region, as hypothesized in the above qPCR experiment, high-throughput sequencing revealed another highly enriched sequence family with unexpected biotinylation codons at the beginning of the coding region. These mutations are likely caused by the omission and deletion of nucleotides during the

chemical synthesis of the library, as well as the small amount of contaminations of incorrect dinucleotides in the degenerative mixture of dinucleotide phosphoramidites.

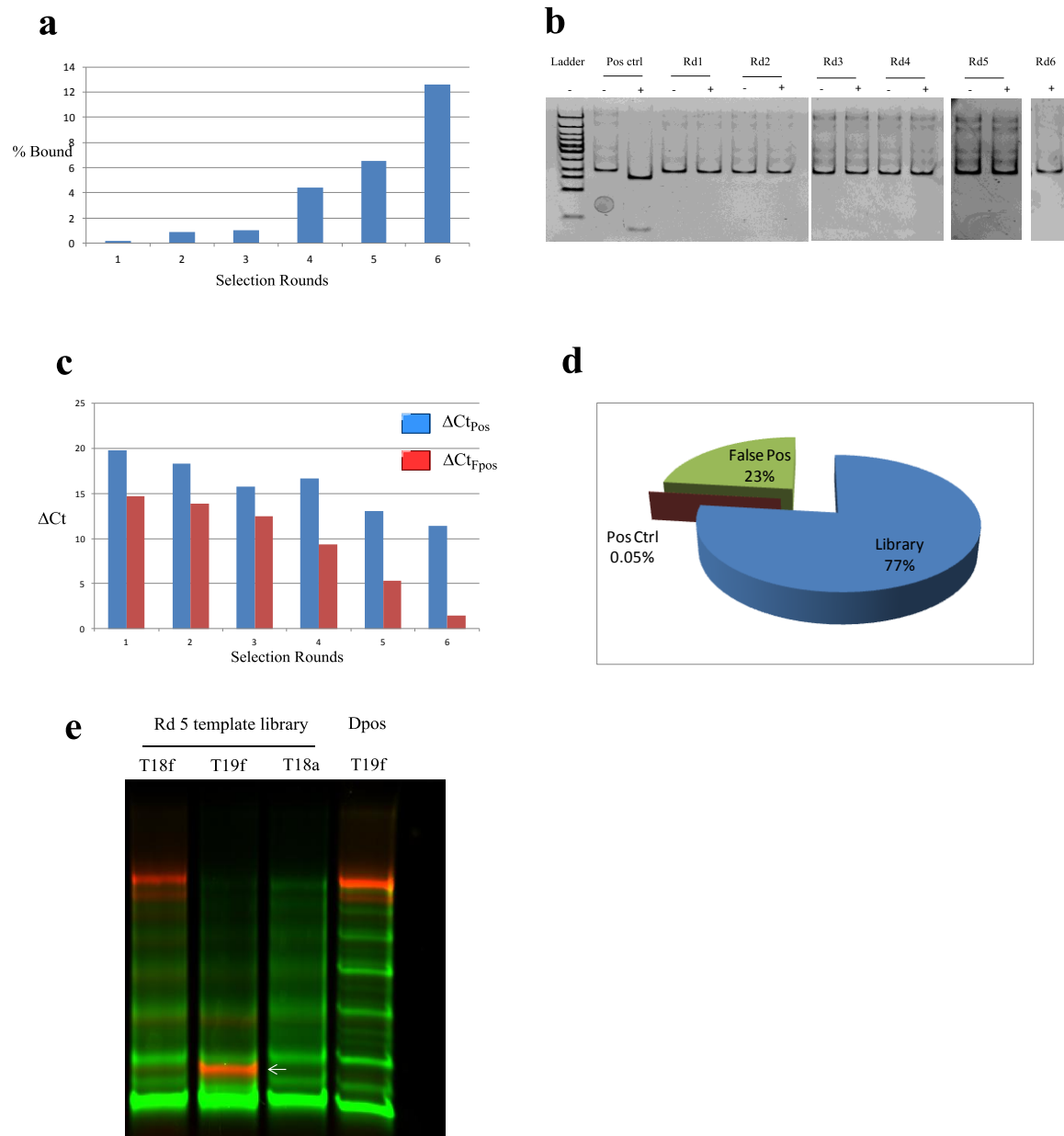


Figure 3.10

Figure 3.10 (Continued)

Figure 3.10. (a) The ratio of bound translation products over all products during the course of mock selection. (b) AseI restriction enzyme digestion of the DNA after each round of PCR amplification. Lane 2 and 3 shows the digestion of the PCR product amplified using **Dpos** as template. (c) qPCR analysis of the enrichments of **Dpos** and a hypothesized mutation that cause unexpected GCATT codon. The ΔC_t values reflect the fraction of the probed DNA within the total library, with the relationship fraction of probed DNA = $2^{-\Delta C_t}$. $\Delta C_{tPos} = C_{tPos} - C_{tLib}$; $\Delta C_{tFpos} = C_{tFpos} - C_{tLib}$; where C_{tPos} , C_{tFpos} , and C_{tLib} are qPCR C_t values using **Rpos**, **Rfpo**, and **Rlib** (5'-CTAGCTAAAAGGCAGG-3', reverse primer of the whole library) as reverse primers. (d) Statistics of the high-throughput sequencing experiments of the DNA template library after six rounds of mock selection. Pos Ctrl indicates the fraction of **Dpos** reads in the entire library. (e) Translation using the DNA template library of Round 5. All reactions used 15 eq. of substrates **Lib1~Lib16** to the DAN template. **T18f**, **T19f**, and **T18a** are the terminators used in each reaction. As a positive control, the translation reaction of lane 4 was performed with **Dpos**, **Lib1~Lib16**, and **T19f**. The white arrow indicates an enriched very short translation product that is labeled with Alexa Fluor 647 on **T19f**, indicating the mutant templates containing unexpected biotinylation codon GCATT has been enriched and resulted the production of very short translation product.

Finally, the ability of enriched mutation in the DNA template library to incorporate biotin groups in the translation product is confirmed by a translation using Round Five template library and terminator **T19f**, a **T19** terminator hybridizing with GCATT codon and labeled with Alexa Fluor 647 (see Method section for the detailed structure of **T19f**) (Figure 3.10e). The Round Five DNA template library installs **T19f** to mainly form very short translation products, instead of full-length polymers produced by successful translation. This result is consistent with the fact that a sequence family with the GCATT mutation at the beginning of the coding region of **LibX**. Moreover, the omission and deletion mutations at the end of the coding region could also cause very

short translation products, as the reading frames have been disrupted in these templates and the terminator installed at the end of the coding region may be able to reach and couple with the hairpin region directly.

3.8 Concluding Remark and Future Directions

We anticipate the enzyme-free DNA-templated translation system developed in this dissertation has the potential to enable the laboratory evolution of a variety of different types of non-nucleic acid sequence-defined synthetic polymers beyond the scope of other enzymatic translation systems. An interesting example of such functional polymers is β -peptide polymers, of which the chemical formulation shares similarities with natural proteins, but the folding patterns are distinctly different. As no enzymatic translation strategy can generate β -peptide polymers with acceptable efficiency, our enzyme-free translation system could enable the evolution of these polymers with novel structures and functions that would otherwise be very difficult to achieve. The design of a genetic code, the preparation of a DNA template library and the macrocyclic substrates containing β -peptide sequences predisposed to polymer folding, and the development of the primer extension and full-length translation product affinity purification strategies will all help the maturation of an *in vitro* selection system of sequence-defined synthetic polymers.

An *in vitro* selection system to evolve sequence-defined synthetic polymer that can bind to a molecular target is shown in Figure 3.11a. Similar to the mock selection described in this dissertation, the *in vitro* selection cycle also involves a translation step, a

post-selection PCR amplification step, and a template re-generation by ligation step. Learning from the lesson of mock selection in which mutations among library members caused the unexpected selection results, we added a full-length product purification strategy and a primer extension strategy in the selection cycle to ensure the quality of the translation products. In addition, we reason that the selection for unknown interactions between library members and a molecular target usually relies on the cooperative binding of multiple building blocks rather than a lone interaction of a single building block with the target, diminishing the chance of a DNA template with one mutated codon to dominate the selection.

Similarly, an in vitro selection system to evolve synthetic polymer catalyst for a bond-forming transformation can also be designed based on the same principles as the binding selection (Figure 3.11b). The functional groups, X and Y, between which the formation of a covalent bond is to be catalyzed by the translated polymer are displayed on the hairpin region of the DNA template. Conjugated to a biotin group, Y is also connected with the DNA hairpin through a cleavable linker. After the translated DNA template library is incubated under a reaction condition that may facilitate the catalysis of a bond-forming transformation by the translation products, the selection criterion will be linker cleavage and subsequent streptavidin capture. Only the translation products that can catalyze the covalent bond formation between X and Y will be able to be captured by streptavidin after linker cleavage. Non-active translation products will lose the biotinylated Y group after cleavage and cannot be captured in principle. We anticipate such a selection system would enable the evolution of robust synthetic polymer catalysts

with the ability to function in the difficult environments that the biopolymer counterparts are subjected to limitations of functional scopes, stability, immunogenicity, and delivery.

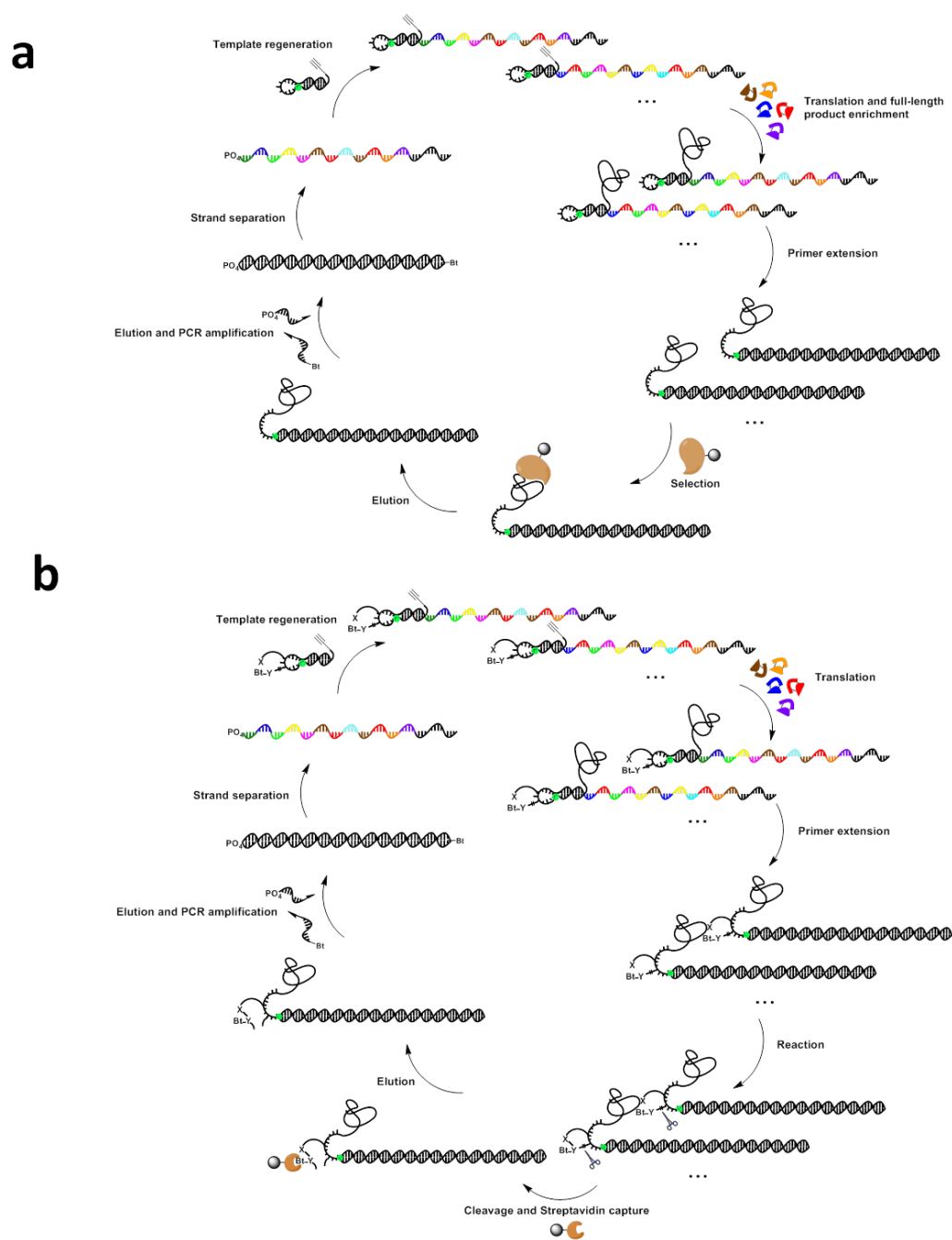


Figure 3.11

Figure 3.11 (Continued)

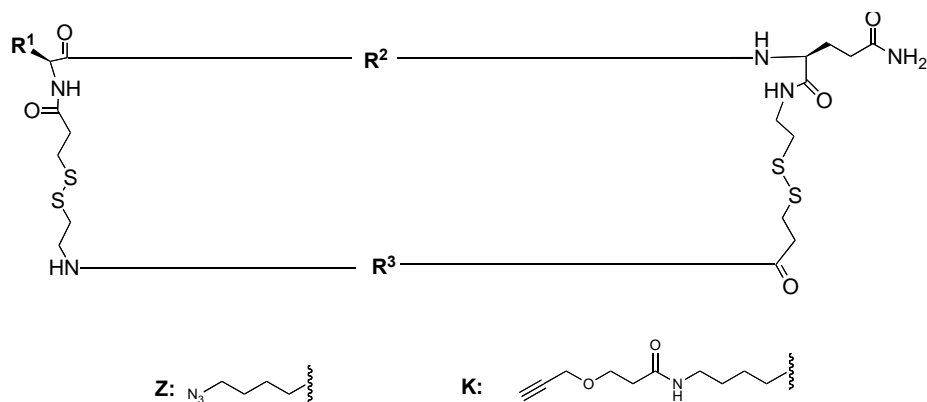
Figure 3.11. (a) An *in vitro* selection system to evolve synthetic polymers that can bind to a molecular target. A DNA template library is translated into sequence-defined synthetic polymers. The resulting translation products undergo full-length product enrichment to purify off the truncations and non-translated templates. The DNA component of the purified translation products are made double-stranded by a DNA polymerase-catalyzed primer extension reaction. The resulting double-stranded DNA-synthetic polymer conjugates undergo selections for binders that can be captured by the beads with the molecular target immobilized. The beads are washed, and the captured translation products are eluted. The elution serves as the template in a subsequent PCR reaction, using a 5'-phosphorylated forward primer and a 5'-biotinylated reverse primer. The PCR products are purified and strand separated. The sense strands with 5'-phosphorylation are saved, and ligated to a DNA hairpin with 5'-hexylnyl modifications by T4 DNA ligase. The resulting DNA can serve as a template of the subsequent round of translation. (b) An *in vitro* selection system to evolve synthetic polymers that can catalyze a bond forming reaction between two reactive partners. A DNA template library with the two reactive groups displayed on the hairpin region is translated into sequence-defined synthetic polymers. One of the reactive groups is biotinylated and connected to the DNA template through a cleavable linkage. The resulting translation products undergo full-length product enrichment to purify off the truncations and non-translated templates. The DNA component of the purified translation products are made double-stranded by a DNA polymerase-catalyzed primer extension reaction. The resulting double-stranded DNA-synthetic polymer conjugates are incubated in a reaction condition that can facilitate the synthetic-polymer catalysis of the covalent bond formation. The translation products are subjected to linker cleavage condition and undergo a size-exclusion step to rid of small molecule species. The resulting library is incubated with streptavidin immobilized beads. The beads are washed, and the captured translation products are eluted. The elution serves as the template in a subsequent PCR reaction, using a 5'-phosphorylated forward primer and a 5'-biotinylated reverse primer. The PCR products are purified and strand separated. The sense strands with 5'-phosphorylation are saved, and ligated to a DNA hairpin with 5'-hexylnyl modifications by T4 DNA ligase. The resulting DNA can serve as a template of the subsequent round of translation.

3.9 Methods

General Methods

Other than the exceptions noted below, all commercially available reagents and solvents were purchased and used without further purification. Dinucleotide phosphoramidites were purchased from Chemgene as degenerative mixtures. DNA template library were synthesized by Integrated DNA Technologies (IDT). All oligonucleotide primers were purchased from IDT. The syntheses of 16 substrates and all terminators for library translation and characterizations followed the same protocol as described in Chapter Two. Minelute spin columns are the corresponding buffers to perform a Reaction Cleanup protocol or a PCR cleanup protocol were purchased from Qiagen. Enzymes were purchased from New England Biolabs. Sanger sequencing was performed by Genewiz. High-throughput DNA sequencing was performed on an in-house Illumina Miseq DNA sequencer, with all sequencing reagents purchased from Illumina. All reagents for oligonucleotide synthesis were purchased from Glen Research.

Table 3.5 Structure of Terminators **T17f**, **T17b**, **T18f**, **T18b**, **T19f**, and **T19b**



Terminator substrate	R^1	R^2	R^3	Post-cleavage modification
T17f	Z	PEG ₂ -Lys-PEG ₂	TGGTA	Alexa Fluor 647 NHS
T17b	K	PEG ₂ -Lys-PEG ₂	TGGTA	Biotin-PEG ₄ -SS-NHS
T18f	Z	PEG ₂ -Lys-PEG ₂	AATCG	Alexa Fluor 647 NHS
T18b	K	PEG ₂ -Lys-PEG ₂	AATCG	Biotin-PEG ₄ -SS-NHS
T19f	Z	PEG ₂ -Lys-PEG ₂	AATGC	Alexa Fluor 647 NHS
T19b	K	PEG ₂ -Lys-PEG ₂	AATGC	Biotin-PEG ₄ -SS-NHS

Sequences of the DNA Templates Used in This Section

Non-standard template functionalities are as follows:

Hex: hexylnyl alkyne modifier.

Cy3: Cy3TM phosphoramidite from Glen Research

Dm1: 5'-/Hex/-GTTTACGACTGGGATCCGTGTCC/Cy3/TCTTCTGGACACG
GATCCCAGTCGTAAACAGTCTCTCAAAACCTTCACAACACTCTACAAACCTTC
CAAAACCTTCACAACACTTCACATACCATTTCCTATAGCCGTTCAG-3'

Ds1: 5'-/Hex/-
GTTTACGACTGGGATCCGTGTCC/Cy3/TCTTCTGGACACGGATCCCAGTCGTAA
ACAACCTTCACAAACCTTCCAAATGTCTCGTAATGCTTCTGA

Ds2: 5'-/Hex/-
GTTTACGACTGGGATCCGTGTCC/Cy3/TCTTCTGGACACGGATCCCAGTCGTAA
ACACATCCTACAACACTCTCAAAGTTCCTGTAAGTCTCTTGA

Df2: 5'-/Hex/-
GTTTACGACTGGGATCCGTGTCC/Cy3/TCTTCTGGACACGGATCCCAGTCGTAA
ACAGTTCCTTGAACACTTCACAAGTCTTCGTAAGTCTCTTGAAGTCTCTGAA
CACTTCCAATACCA-3'

Df3: 5'-/Hex/-
GTTTACGACTGGGATCCGTGTCC/Cy3/TCTTCTGGACACGGATCCCAGTCGTAA
ACATGTCCTCAAACATCCTACAAGTCTCTTGAATGCTCTACAATGCTCTCAAA
CATCCTTGATACCA-3'

Df4: 5'-/Hex/-
GTTTACGACTGGGATCCGTGTCC/Cy3/TCTTCTGGACACGGATCCCAGTCGTAA
ACACATCTCTGAATGTCCTTGAAACCTCTACAAACCTCTTGAAACTCTCTGAA
GTTCTCGTATACCA-3'

Df6: 5'-/Hex/-
GTTTACGACTGGGATCCGTGTCC/Cy3/TCTTCTGGACACGGATCCCAGTCGTAA
ACAACCTTCCAAAGTCTCTACAACACTCTTGAAGTTCCTCAAAGTCTTCTGAA
CACTTCTGATACCA-3'

Df7: 5'-/Hex/-
GTTTACGACTGGGATCCGTGTCC/Cy3/TCTTCTGGACACGGATCCCAGTCGTAA
ACAACCTTCACAAACCTCTTGAATGTCCTCAAAGTTCCTCAAAGTCTCTCAAA
ACTCCTTGATACCA-3'

Df8: 5'-/Hex/-
GTTTACGACTGGGATCCGTGTCC/Cy3/TCTTCTGGACACGGATCCCAGTCGTAA
ACAGTTCCTGTAACATCCTCAAAACCTTCCAAAGTCTCTGTAACACTCTACAA
ACTCTCCAATACCA-3'

Df9: 5'-/Hex/-
GTTTACGACTGGGATCCGTGTCC/Cy3/TCTTCTGGACACGGATCCCAGTCGTAA
ACATGTCCTTGAATGTCCTTGAATGTCCTGTAAGTTCCTTGAAACTCCTCAAA
GTCTCTGTATACCA-3'

Df15: 5' - /Hex/-

GTTTACGACTGGGATCCGTGTCC/Cy3/TCTTCTGGACACGGATCCCAGTCGTAA
ACAACCTCTACAAACCTCTCAAACTCTCACAACATCTCACAACTCTCCAA
AACCTCTCAATACCA-3'

Dl1: 5' -/Hex/-

CCTAGCCCGAAG/Cy3/CCCTTGAACCTCCTCCTTCGGGCTAGGAATCCATACC
AATCCATACCAATCCCATCAGAAGCCAGAGGCGAACGAATG-3'

Sample Preparation for Illumina High-Throuput Sequencing of the DNA Library

A. Two Stage PCR Amplification to Install the Sequencing Adapters.

One microliter of 3.6 nM DNA template from the DNA template library, **LibX**, was combined with 50 µL Q5 hot-start DNA polymerase master mix, 100 pmol of forward primer (5'-
ACACTCTTTCCTACACGACGCTCTTCCGATCTNNNNGATCCCAGTCGTAAAC
-3'), 100 pmol of reverse primer (5'-
TGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNCTAGCTAAAAGGCAGG-3')
and water to a total volume of 100 µL. The Ns in the primer sequences are random nucleotides coupled as a degenerative mixture of all four bases. Twenty-five microliters of such mixture is added with 1.25 µL of 10x SYBR Green I fluorescent dye to perform qPCR. The cycles of PCR to reach fluorescent saturation on the qPCR were used to perform normal PCR for the remaining PCR mixture. PCR was performed using the following temperature schedule: 98 °C for 30 s, followed by iterative cycles of [98 °C for 10s, 68 °C for 30 s, 72 °C 10 s]. After PCR, the resulting materials were processed using a Qiagen Minielute PCR purification kit and directly use as a template for the second

PCR. 1/1000th of the PCR product from the previous reaction were combined with 50 µL Q5 hot-start DNA polymerase master mix, 100 pmol of Illadapterfwd2-B (5'-AATGATACGGCGACCACCGAGATCTACAC TCCGGAGA ACACTCTTTCCCTACACGAC-3') and 100 pmol of PE_REV-20B (5'-CAAGCAGAAGACGGCATACGAGATGGCCACAAGTGACTGGAGTTCAGACGT GTGCT-3') and water to a total volume of 100 µL. Twenty-five microliters of such mixture is added with 1.25 µL of 10x SYBR Green I fluorescent dye to perform qPCR. The cycles of PCR to reach fluorescent saturation on the qPCR were used to perform normal PCR for the remaining PCR mixture. PCR was performed using the following temperature schedule: 98 °C for 30 s, followed by iterative cycles of [98 °C for 10s, 68 °C for 30 s, 72 °C 10 s]. After PCR, the resulting materials were processed using a Qiagen Minielute PCR purification kit and subsequently purified on a native PAGE. The DNA eluted from the native PAGE were further purified a Qiagen Minielute Enzymatic Reaction Cleanup purification kit.

B. Picogreen Quantification of DNA

Prepare 1.5 mL 1x Picogreen by diluting from 7.5 µL of 200x Picogreen stock solution using 1x TE buffer supplied by the Picogreen quantification kit. Prepare 150 µL of 2 µg/mL standard DNA sample by diluting from 100 µg/mL DNA in the kit. Prepare a series of five samples with the concentration of standard DNA to be 2 µg/mL, 0.2 µg/mL, 0.02 µg/mL, 0.002 µg/mL and 0 µg/mL (straight TE), each 100 µL. Add 1.5 µL of purified PCR product from last step into 148.5 µL of TE buffer and mix well. Add 75 µL

of each standard and sample solution into a well of a 96-well plate. Mix with 75 μ L of 1x picogreen. Mix well and store in dark for 3 min. Read the fluorescence on the plate reader (Excitation 480 nm, emission 520 nm). Draw a standard curve using the DNA standard and back calculate the concentration of the sample PCR product.

C. KAPA qPCR quantification of DNA

Based on Picogreen quantification result, dilute the PCR products into a “4nM” solution. Make dilution series of 1/250, 1/500, 1/1000, 1/2000, 1/4000 of the “4nM” DNA sample. Add 12 μ L of Kappa Biosciences 2x qPCR mix, 4 μ L of water, and 4uL of the dilution series of “4nM” DNA sample. In a separate PCR strip tube, add a 12 μ L of Kappa Biosciences 2x qPCR mix, 4 μ L of water, and 4uL of Kappa Kit Stand 1, Stand 2, Stand 3, Stand 4, Stand 5, and Stand 6, respectively. PCR was performed using the following temperature schedule: 95 °C for 5 min, followed by 30 cycles of [95 °C for 30s, and then 60 °C for 45 s]. The cycles of PCR to reach fluorescent saturation on the qPCR were used to perform a separate PCR, of which the products were analyzed by a native PAGE. Draw a standard curve using the DNA standard, back calculate the concentration of the sample PCR product and make adjustment of concentration accordingly to reach exactly 4 nM concentration. The 4 nM DNA sample were directly subjected to Illumina Sequencing following a standard sequencing protocol.

Optimized DNA-Templated Library Translation Protocol

Translation reactions were prepared by combining 50 μ L of degassed [0.1 M aqueous HEPES, 0.2 M NaCl] pH 8.0 buffer, 10 pmol of DNA template, 150 pmol of each **Lib1~Lib16** substrate (15 eq. per DNA template), 100 pmol of the terminator substrate, 1 μ mol tris(hydroxypropyltriazolyl)amine, THPTA,²³ 1 μ mol CHAPS, and water to a total volume of 100 μ L. The reaction was heated to 95 °C and slowly cooled to 5 °C, at a ramp rate of 0.1 °C/10s. The CuAAC reaction was initiated by the addition of 2 μ L 50 mM CuSO₄, followed by the addition of 2 μ L 0.5 M sodium ascorbate. The reaction mixture was maintained at 4 °C overnight. After incubation, the reaction was purified using a Qiagen Minelute Enzymatic Reaction Cleanup kit, eluting into 10 μ L 0.1 M Ammonium bicarbonate (ABC) buffer. The disulfide linkages between the polymer product and the adapters were cleaved by addition of 1 μ L 0.1 M TCEP solution. The cleavage reaction was incubated for 45 min before adding 1 μ L 0.25 M IAA to alkylate the free thiols. The alkylation was incubated for 30 min before quenching by the addition of 1 μ L 1 M DTT. The quenched solution was analyzed by 10 % denaturing PAGE. The gel was stained with ethidium bromide and analyzed by UV illumination and densitometry.

Affinity Purification of the Full-Length Translation Product

Translation reactions were prepared by combining 50 μ L of degassed [0.1 M aqueous HEPES, 0.2 M NaCl] pH 8.0 buffer, 10 pmol of DNA template, 150 pmol of each **Lib1~Lib16** substrate (15 eq. per DNA template), 100 pmol of biotinylated terminators, 1 μ mol THPTA, 1 μ mol CHAPS, and water to a total volume of 100 μ L.

The reaction was heated to 95 °C and slowly cooled to 5 °C, at a ramp rate of 0.1 °C/10s. The CuAAC reaction was initiated by the addition of 2 µL 50 mM CuSO₄, followed by the addition of 2 µL 0.5 M sodium ascorbate. The reaction mixture was maintained at 4 °C overnight. After incubation, the reaction was purified using a Qiagen Minelute Enzymatic Reaction Cleanup kit, eluting into 10 µL 0.1 M ABC buffer. The purified translation reaction was added into Streptavidin magnetic particles washed and then suspended in 10 µL [10 mM Tris, 2 M NaCl, 1 mM EDTA, pH 8.0] buffer. The suspension was incubated at room temperature for 30 min before the beads were pulled down by a magnet. The beads were washed three times, and the bound biotinylated translation products were eluted by the treatment of the beads with 10 mM TCEP in 0.1 M ABC for 45 min. The beads were subsequently pulled down and the supernatant was added 1 µL 0.25 M IAA to alkylate the free thiols. The alkylation was incubated for 30 min before quenching by the addition of 1 µL 1 M DTT. The quenched solution was analyzed by 10 % denaturing PAGE. The gel was stained with ethidium bromide and analyzed by UV illumination and densitometry.

Primer Extension to Synthesize the Complementary DNA for the DNA Template after Translation

Twenty microliters of Translation products after adapter cleavage were first desalted by a Sephadex minicolumn (Princeton Separations). The resulted solution was added with 3 µL of NEBuffer 2 10x, 1 µL of 10 µM reverse primer **RLib** (5'-CTAGCTAAAAGGCAGG-3'), 1.5 µL of 2 mM dNTP, and water to a total volume of 30

μL. The reaction was heated to 70 °C and slowly cooled to 37 °C, at a ramp rate of 0.1 °C/s. The primer extension reaction was initiated by adding 5 units of Klenow Fragment (3'→5' exo-) polymerase. The reaction was kept on a PCR thermal cycler at constant temperature of 37 °C for 1 hour. Subsequently the reaction was purified by a Qiagen Minelute Enzymatic Reaction Cleanup kit and eluted into 10 μL EB buffer supplied with the kit.

Complete Cycle of Translation, PCR, Strand Separation, and Re-Translation Using a Single DNA Template

A. Ligation to generate the starting template D4

7.5 nmol of the reverse primer **RP** was mixed with 3 nmol of the coding strand (5'-Phos-GGCTAGGAATCCATACCAATCCATACCAATCCCATCAGAAGCCACATGCGAACGAATG-3') in T4 DNA ligase buffer. The mixture was heated to 65 °C and cooled to 16 °C at a rate of 0.1 °C per min. At 16 °C, 5,000 units of T4 DNA ligase were added, and the reaction was maintained at 16 °C for 16 h. The DNA in the mixture was precipitated with cold ethanol. The ligated template was purified by denaturing PAGE and desalted.

B. DNA-templated translation and gel purification

DNA-templated oligomerization was performed by mixing 0.4 mL of degassed 0.1 M HEPES pH 8.0 buffer, 10 μL of 4 M NaCl, 0.1 nmol of DNA template **D4**, 120 pmol **22**, 80 pmol **23**, 40 pmol **T7**, and 50 μL of 0.1 M Cu(I) ligand THPTA in a total volume of 0.5 mL. The reaction was heated to 95 °C and slowly cooled to 5 °C. CuAAC was initiated by the addition of 10 μL 50 mM CuSO₄, followed by the addition

of 10 μL 0.5 M sodium ascorbate. The water used to dissolve CuSO_4 and sodium ascorbate was previously deoxygenated by blowing nitrogen gas through the water for 20 min. The reaction mixture was incubated at 5 $^\circ\text{C}$ overnight. After incubation, the reaction was desalted using a Princeton Separation minicolumn. The sample was lyophilized. The resulting materials were dissolved in 9 μL of 0.1 M ammonium bicarbonate buffer. 1 mL 0.2 M DTT was added and the sample was incubated at 70 $^\circ\text{C}$ for 10 min in darkness. The sample was then cooled to room temperature, and 2 μL of 0.25 M iodoacetamide was added. The alkylation reaction was incubated in darkness for 30 min, and 1.4 μL DTT was added to quench the reaction. The resulting sample was subjected to 10 % denaturing PAGE. The product band was visualized by UV shadowing and excised. The excised gel was homogenized and the translation products were eluted with 400 μL of 10 mM Tris-EDTA pH 7.4 buffer overnight at 37 $^\circ\text{C}$ under constant shaking. After filtration to remove the gel particles, the filtrate was concentrated to 100 μL and desalted by a Princeton Separation minicolumn.

C. PCR amplification and strand separation to re-generate the template

1/50,000th of the gel-purified translation product was combined with 0.8 mL Q5 hot-start DNA polymerase master mix, 1.6 nmol **RP**, 1.6 nmol **FP** and water to a total volume of 1.6 mL. PCR was performed using the following temperature schedule: 98 $^\circ\text{C}$ for 30 s, followed by 30 cycles of [98 $^\circ\text{C}$ for 10s, 68 $^\circ\text{C}$ for 30 s, 72 $^\circ\text{C}$ 10 s]. After PCR, the resulting materials were processed using a Qiagen Minielute PCR purification kit and then purified on a 10% native PAGE. After elution of the purified DNA from gel and

desalting using a Princeton Separation minicolumn, the sample was incubated with 30 μ L of 10 mg/mL Invitrogen MyOne C1 streptavidin-immobilized magnetic beads for 30 min. The beads were washed with 1 \times TBS buffer and incubated with 0.15 M NaOH for 10 min. The eluant contained the single-stranded “sense” template for subsequent re-translation.

D. Re-translation using the template generated by PCR and strand separation

Re-translation was performed using the same conditions used in the translation reaction, except on a 5-pmol scale. The reaction was desalted with a Princeton Separation minicolumn and lyophilized. The sample was re-dissolved in 10 μ L of 50% aqueous formamide and subjected to 10% native PAGE. 5 pmol aliquots of each of the previous steps of the cycle were also run side-by-side with this sample. The gel was stained by ethidium bromide.

Mock Selection Protocol

Day 1:

1. In a PCR tube was added the DNA library (**LibX**, 10 pmol), substrates **Lib1~Lib16**, 150 pmol each substrate, terminator **T18a** (100 pmol), Terminator **T19b** (50 pmol), THPTA (10 μ L of 0.1 M solution), CHAPS (10 μ L of 0.1 M solution), and [0.1 M HEPES, 0.2 M NaCl, pH 8] degassed buffer up to 100 μ L volume. The reaction was first heated to 95 $^{\circ}$ C for 3 min, and then slowly cooled down to 5 $^{\circ}$ C at a ramp rate of

0.1 °C/10s. The annealed reaction was then transferred to ice, added CuSO₄ (2 µL of 50 mM solution in degassed water) and sodium ascorbate (2 µL of 0.5 M solution in degassed water), and moved to the cold room (4 °C constant temperature). The polymerization reaction was placed on a shaker in the cold room overnight, protected from light.

Day 2:

2. After overnight incubation, the polymerization reaction was purified by a Qiagen Minelute spin column, using Enzymatic Reaction Cleanup protocol (briefly, add ERC buffer 300 µL with 10 µL of 3 M pH 5.2 sodium acetate; mix and spin at 13,000 rpm for 1 min; add 750 µL PE buffer; spin; discard flow through; spin; add 10 µL 0.1 M ammonium bicarbonate (ABC) buffer; wait for 5 min and spin).

3. In a 1.5 mL centrifuge tube was added 20 µL Dynabeads Streptavidin C1 magnetic beads (Invitrogen, 65001). The beads were washed according to the instruction manual [3 × 30 µL 1x B&W buffer(5 mM Tris, 0.5 mM EDTA, 1 M NaCl)], incubated with blocking buffer (1 mg/mL Salmon Sperm DNA in 1 x B&W buffer) and washed by 1 x B&W buffer, and then resuspended in 10 µL 2x B&W buffer. Subsequently the purified translation reaction (10 µL in 0.1 M ABC buffer) was added. The mixture was mixed well and placed on a rotary for 30 min under room temperature. The magnetic beads were pulled down on a magnet, and the supernatant was pipetted out carefully and saved as flow through (FT) after a desalting step using a Princeton Separations Column (CS-901, Princeton Separations). The magnetic beads with biotinylated translation

product bound were washed [$3 \times 30 \mu\text{L}$ 1x B&W buffer plus 0.1 % Tween 20, $3 \times (0.1 \text{ M Tris, } 8\text{M urea, pH } 8)$ buffer, $3 \times (20\text{mM NaOH, } 1 \text{ mM EDTA})$ buffer, 2 x water, and 1x 0.1 M ABC]. The magnetic beads were resuspended in $20 \mu\text{L}$ freshly prepared (10 mM TCEP, 0.1 M ABC, pH 8) solution and placed on a rotary for 45 min under room temperature, protected from light. The magnetic beads were pulled down on a magnet, and the supernatant was carefully pipetted out and saved as the selection (SEL). All TCEP treated samples including SEL and FT, were added freshly prepared 0.25 M iodoacetamide aqueous solution for $1 \mu\text{L}/10 \mu\text{L}$ TCEP containing solution and incubated for 30 min, protected from light. The alkylated samples were then added 1 M DTT for $1 \mu\text{L}/10 \mu\text{L}$ sample and incubated for another 30 min, protected from light. The samples were added water up to the total volume $>20 \mu\text{L}$, and desalted by Princeton Separation columns.

4. For the initial rounds of selection, save two thirds of SEL for future characterization. To the remaining SEL, add 1 nmol of PCR primers: RP-2 and FP-2, and $500 \mu\text{L}$ of Q5 hot-start 2x mastermix and water up to the total volume of $1000 \mu\text{L}$. Take $12.5 \mu\text{L}$ of such mixture and add $12.5 \mu\text{L}$ Q5 hot-start 2x mastermix and $1.25 \mu\text{L}$ 10 x SYBR Green I dye. The qPCR samples for FT were prepared according to the following table:

Table 3.6. Reagents for qPCR of Flow Through (FT)

Reagents	Volume
10 μ M FP-2	2.5 μ L
10 μ M RP-2	2.5 μ L
Q5 2x master mix	12.5 μ L
10x SYBR Green I	1.25 μ L
Template	1 μ L of 1:100 diluted FT
Water	5.25 μ L

The SYBR Green I dye containing samples were subjected to qPCR to determine the relative concentration of biotinylated translation product with respect to non-biotinylated ones and determine the optimal cycling numbers of the PCR to amplify the DNA of SEL. Add 400 μ L Q5 hot-start master mix to the rest of PCR sample. The PCR was programmed accordingly: 98 °C for 30 s, followed by corresponding numbers of cycles of [98 °C 10 s, 56 °C 30 s, 72 °C 10s]. The crude PCR material was purified with Qiagen Minelute spin columns using PCR cleanup protocol (briefly, add PB buffer 5 times of volume of PCR crude sample with 10 μ L of 3 M pH 5.2 sodium acetate; mix and spin at 13,000 rpm for 1 min; add 750 μ L PE buffer; spin; discard flow through; spin; add 10 μ L EB buffer; wait for 5 min and spin). Quantify the purified PCR products.

Day 3:

5. In a 1.5 mL centrifuge tube was added 110 μ L Dynabeads Streptavidin C1 magnetic beads. The beads were washed according to the instruction manual [$3 \times 150 \mu$ L 1x B&W buffer(5 mM Tris, 0.5 mM EDTA, 1 M NaCl)], and then resuspended in 80 μ L

2x B&W buffer. Subsequently the purified PCR product (80 μ L in EB buffer) was added. The mixture was mixed well and placed on a rotary for 30 min under room temperature. The magnetic beads were pulled down on a magnet, and the supernatant was pipetted out carefully and saved. The beads were then washed three times with 200 μ L of 1x B&W buffer. 40 μ L of freshly prepared 150 mM NaOH was added to the beads, and the suspension was incubated at room temperature on a rotary for 15 minutes. The liberated strand was separated from the magnetic beads by use of a magnet, and the separated strand was immediately added 10 μ L 3M NaOAc pH 5.2 buffer, and desalted by gel filtration using a Princeton Separations column equilibrated with water. Measure the concentration of the resulting single-stranded DNA template library using Nanodrop.

6. In a PCR tube was added the single-stranded enriched DNA template library, 400 pmol of the hairpin DNA T3-Hex, 10 μ L of 10x T4 DNA ligase buffer, and water up to total volume of 100 μ L. Heat the reaction to 70 $^{\circ}$ C and slowly cool down to 16 $^{\circ}$ C at a ramp rate of 0.1 $^{\circ}$ C/s. Add T4 DNA ligase 1 μ L (400 U) and incubate at 16 $^{\circ}$ C for 1 h. Ethanol precipitate the reaction mixture by adding 10 μ L of pH 5.2 3M NaOAc and 256 μ L of cold neat ethanol. Keep the cold mixture on dry ice for 20 min and spin under 15,000 rpm for 30 min in the cold room. Dry the pellet under reduced pressure. Add 10 μ L of water and 10 μ L of formamide to dissolve the crude sample and load it on a 10 % TBE-urea denaturing PAGE. Run the gel at 200 V for 1 h, and excise the band corresponding to the desired product. Spin-extrusion the gel to break it into fine particles, and add 1x TE buffer 200 μ L to elute the DNA out overnight on 37 $^{\circ}$ C shaker.

Day 4:

7. Filter the gel particles using a 0.22 μm pore size centrifuge filter. The supernatant was added 20 μL of pH 5.2 3 M NaOAc buffer, and 510 μL cold neat ethanol to precipitate the DNA. Keep the cold mixture on dry ice for 20 min and spin under 15,000 rpm for 30 min in the cold room. Wash the pellet once with cold 70% aqueous ethanol and dry it under reduced pressure. Redissolve the DNA in 20 μL of water and purify with a Princeton Separations column. Quantify the DNA concentration. This would give the starting point of another round of selection.

MATLAB Program for Searching for Codons in Genetic Code Design

```
clear all;

%FORCESET = [0 0] gives the best in regard to minimizing 5bp off-targets
FORCESET = [3 4];

HIST_RANGE = [-20 10];
POL_LEN=12;
N_LIB=1000;
TEMPERATURE='25';
NACONC='.2';
MGCONC='0';
NUM_C=16;
GU_WOBBLE=1;

possible(1).first= {'AC'; 'CA'};
```

```

possible(1).second= {'ACT';'ATC';'CAT';'TAC';'CTA';'TTG'};
possible(2).second= {'AC';'TG'};
possible(2).first= {'ATC';'ACT';'CAT';'GTT';'CTA';'TGT'};

possible(3).first= {'TC';'CT'};
possible(3).second= {'ACA';'CAA';'GTA';'TGA'};
possible(4).second= {'TC';'CT'};
possible(4).first= {'AAC';'ATG';'ACA';'AGT'};

possible(5).first= {'TA';'AT'};
possible(5).second= {'ACC';'CCA';'CTG';'TGC'};
possible(6).second= {'TA';'AT'};
possible(6).first= {'ACC';'CCA';'CTG';'GCT'};

for C = 1:length(possible)
    for A = 1:length(possible(C).first)
        for B = 1:length(possible(C).second)
            apos(C).c(length(possible(C).second)*(A-
1)+B).code=[char(possible(C).first(A)) char(possible(C).second(B))];
        end
    end
end

end

boffscore5=(NUM_C^3)/4;
boffscore4=(NUM_C^3)/4;
boffscore3=(NUM_C^3)/4;

```

```

foundno5=0;
foundno4=0;
foundno3=0;
totcount=1;
for C1 = 1:length(possible)
    for C2 = C1+1:length(possible)
        totscore(totcount,1:4)=0;
        totcount=totcount+1;
        if C1 == C2

        else

            for B1 = 1:NUM_C/2
                for B2 = 1:NUM_C/2
                    for B = 1:NUM_C/2
                        codonscore1(B1,B2,B)=Jiastrcmp([apos(C1).c(B1).code
apos(C2).c(B2).code],apos(C1).c(B).code);
                        codonscore2(B1,B2,B)=Jiastrcmp([apos(C2).c(B2).code
apos(C1).c(B1).code],apos(C1).c(B).code);
                        codonscore3(B1,B2,B)=Jiastrcmp([apos(C1).c(B1).code
apos(C2).c(B2).code],apos(C2).c(B).code);
                        codonscore4(B1,B2,B)=Jiastrcmp([apos(C2).c(B2).code
apos(C1).c(B1).code],apos(C2).c(B).code);
                    end
                end
            end

            offscore5=length(find(codonscore1 >=
5))+length(find(codonscore2 >= 5))+length(find(codonscore3 >=
5))+length(find(codonscore4 >= 5));

            offscore4=length(find(codonscore1 >=
4))+length(find(codonscore2 >= 4))+length(find(codonscore3 >=
4))+length(find(codonscore4 >= 4));

```

```

        offscore3=length(find(codonscore1 >=
3))+length(find(codonscore2 >= 3))+length(find(codonscore3 >=
3))+length(find(codonscore4 >= 3));

    totscore(totcount-1,1:4) = [C1 C2 offscore5 offscore4 ];

    if offscore5 <= boffscore5
        foundno5=foundno5+1;
        boffscore5=offscore5;
        if boffscore4 == (NUM_C^3)/4
            bcodonp=[C1 C2];
        end
        if offscore4 <= boffscore4

            bcodonp=[C1 C2];

            foundno4=foundno4+1;
            boffscore4=offscore4;
            boffscore3=offscore3;
        end
    end
end
end
end
end

display('          Sets      5bp  4bp  (Out-of-frame annealing) ')
disp(totscore)

if FORCESET(1) == 0

```



```

        display(['Best codon sets: ' num2str(bcodonp(2)) ' '
num2str(bcodonp(1))])

    else

        display(['Forced codon sets: ' num2str(FORCESET(2)) ' '
num2str(FORCESET(1))])

        bcodonp(1)=FORCESET(1);

        bcodonp(2)=FORCESET(2);

    end

%display(['Out-of-frame annealing 5bp: ' num2str(boffscore5) ' 4bp: '
num2str(boffscore4) ' 3bp: ' num2str(boffscore3) ])

display('Female      Male')

for B1 = 1:NUM_C/2

    disp([num2str(2*B1-1) ' ' apos(bcodonp(2)).c(B1).code ' '
num2str(2*B1) ' ' apos(bcodonp(1)).c(B1).code])

        bcodon(2*B1-1).code=char(apos(bcodonp(2)).c(B1).code);

        bcod(B1).female=char(apos(bcodonp(2)).c(B1).code);

        bcodon(B1*2).code=char(apos(bcodonp(1)).c(B1).code);

        bcod(B1).male=char(apos(bcodonp(1)).c(B1).code);

end

display(' ')

display('5bp Out-of-frame annealing ')

for B1 = 1:NUM_C/2

    for B2 = 1:NUM_C/2

        for B3 = 1:NUM_C

            for rev = 1:2

                if rev == 1

                    s=[ bcodon(2*B1-1).code bcodon(B2*2).code];

                elseif rev == 2

```

```

                                s=[ bcodon(B2*2).code bcodon(2*B1-1).code];
                                end

c=bcodon(B3).code;
count=1;
bmm(1)=0;
for A = 1:length(s)-length(c)-1
    mm = 0;
    for B = 1:length(c)
        if s(A+B) == c(B)
            mm =mm+1;
        end
    end
    if bmm(1) <= mm
        bmm(1)=mm;
    end
end
for A = 1:length(s)-length(c)-1
    mm = 0;

    for B = 1:length(c)

        if s(A+B) == c(B)

            mm =mm+1;
        end
    end

end

```



```

        for rev = 1:2
            if rev == 1
                s=[ bcodon(2*B1-1).code bcodon(B2*2).code];
            elseif rev == 2
                s=[ bcodon(B2*2).code bcodon(2*B1-1).code];
            end
        end

c=bcodon(B3).code;
count=1;
bmm(1)=0;
for A = 1:length(s)-length(c)-1
    mm = 0;
    for B = 1:length(c)
        if s(A+B) == c(B)
            mm =mm+1;
        end
    end
    if bmm(1) <= mm
        bmm(1)=mm;
    end
end
for A = 1:length(s)-length(c)-1
    mm = 0;
    cend = 0;
    for B = 1:length(c)

        if s(A+B) == c(B)

            mm =mm+1;

```

```

else
    if B == 1 || B == length(c)
        cend=1;
    end
end

end

if bmm(1) == mm && mm >= 4 && cend == 1
    numcend=numcend+1;
    clear blank;
    blank(1:A+length(num2str(2*B1-1))+length(num2str(2*B2))+2-length(num2str(B3)))=' ';
    disp([num2str(B3) blank c]);
    if rev == 1

        disp([num2str(2*B1-1) '-' num2str(2*B2) ' ' s]);
    elseif rev == 2
        disp([num2str(2*B2) '-' num2str(2*B1-1) ' ' s]);
    end
    count4=count4+1;
    disp(blank)
end

end

end

end

end

end

end

display('4bp Out-of-frame annealing with G-T Mismatch ')

```

```

count4=0;
numcend=0;

for B1 = 1:NUM_C/2
    for B2 = 1:NUM_C/2
        for B3 = 1:NUM_C
            for rev = 1:2
                if rev == 1
                    s=[ bcodon(2*B1-1).code bcodon(B2*2).code];
                elseif rev == 2
                    s=[ bcodon(B2*2).code bcodon(2*B1-1).code];
                end
            end

            c=bcodon(B3).code;
            count=1;
            bmm(1)=0;
            for A = 1:length(s)-length(c)-1
                mm = 0;
                for B = 1:length(c)
                    if s(A+B) == c(B)
                        mm =mm+1;
                    end
                end
                if bmm(1) <= mm
                    bmm(1)=mm;
                end
            end

            for A = 1:length(s)-length(c)-1
                mm = 0;
                GU = 0;
                for B = 1:length(c)

```

```

    if s(A+B) == c(B)

        mm =mm+1;

    elseif s(A+B) == 'G' && c(B) == 'A'

        GU = GU+1;

    elseif s(A+B) == 'T' && c(B) == 'C'

        GU = GU+1;

    end

end

end

if mm >= 4 && GU == 1

    clear blank;

    blank(1:A+length(num2str(2*B1-1))+length(num2str(2*B2))+2-length(num2str(B3)))=' ';

    disp([num2str(B3) blank c]);

    if rev == 1

        disp([num2str(2*B1-1) '-' num2str(2*B2) ' ' s]);

    elseif rev == 2

        disp([num2str(2*B2) '-' num2str(2*B1-1) ' ' s]);

    end

    count4=count4+1;

    disp(blank)

end

end

end

```

```

        end

    end

end

filename='tempLib.txt';
unix(['rm ' filename]);
unix(['rm blah']);
unix(['rm ' filename '.dG']);

fileID = fopen(filename,'w');

for n = 1:N_LIB
    clear Libm;
    clear name;
    name='';
    Libm='';
    rn=randsample(length(bcod),POL_LEN,true);

    for A = 1:POL_LEN/2
        Libm=strcat(Libm,[bcod(rn(2*A-1)).female bcod(rn(2*A)).male]);
        name=strcat(name,[num2str(rn(2*A-1)*2-1) '-' num2str(rn(2*A)*2)
        '-']);
    end
    fprintf(fileID,'%s\n',name);
    fprintf(fileID,'%s\n',Libm);
end

fclose(fileID);

```



```
command = ['-t' TEMPERATURE ' -T' TEMPERATURE ' -N' NACONC ' -M' MGCONC  
' -nDNA '];
```

```
unix(['./hybrid-ss-min ' command filename ' > blah' ]);
```

```
unix('awk ' NR % 2 ==0 {print; }' tempLib.txt > temptempLib.txt');
```

```
unix('echo " " > temptempLib2.txt');
```

```
unix('cat temptempLib.txt >> temptempLib2.txt');
```

```
unix(['paste -d"\t" ' filename '.dG' ' temptempLib2.txt | cut -f2,4 >  
temptempLib3.txt']);
```

```
unix('sort -n temptempLib3.txt > Final.txt');
```

```
%display('Library of Random Codons:')
```

```
%unix(['cat ' filename]);
```

```
%display(' ')
```

```
data=importdata( [filename '.dG']);
```

```
display('Most Structured DNAs from Random Library:')
```

```
[Y,I]=sort(data.data(:,2));
```

```
for A = 1:5
```

```
tempnum1=num2str(round(I(A)*2-1));
```

```
tempnum2=num2str(round(I(A)*2));
```

```
unix(['sed -n ' tempnum1 ', ' tempnum2 'p' ' ' filename]);
```

```

display(['dG = ' num2str(data.data(I(A),2))]);

%unix(['sed -n ''' num2str(round(I(1)+1)) ',' num2str(round(I(1)+1))
'p''' ' ' filename '.dG']);

end

set(gca,'FontSize',14);
hist(data.data(find(data.data(:,2)<200),2))
xlabel('Free Energy DeltaG (kcal/mol) ');
ylabel('Sequence Count ');

W=find(data.data(:,2)>=200);
if isempty(W) == 0
display('Weird Structured DNAs from Random Library:')

for A = 1:length(W)

tempnum1=num2str(round(W(A)*2-1));
tempnum2=num2str(round(W(A)*2));
unix(['sed -n ''' tempnum1 ',' tempnum2 'p''' ' ' filename]);

display(['dG = ' num2str(data.data(W(A),2))]);

%unix(['sed -n ''' num2str(round(I(1)+1)) ',' num2str(round(I(1)+1))
'p''' ' ' filename '.dG']);

end

end

%}

```

MATLAB Program for Analyzing the Codon Distribution of the High-Throughput Sequencing Results

```

tic

clear tableOfResults

countTotal = 0;

invalidData = 0;

HasN = 0;

InDel = 0;

PrimMut = 0;

tableOfResults(20,20)=0;

fid = fopen('Jia-PerfectLibraryMembers_2.txt', 'at');

name = 'Jia-library_2.txt';


for x = 1:200

    z = x*200000

    y = z-200000+1

    toc

    [Header, Sequence] = fastqread('2_S2_L001_R1_001.fastq', 'Blockread', [y
z]);

    rawData = char(Sequence); %converts cell array
    to character array, putting each letter in its own column

    clear Sequence

    clear Header

    [amountOfData lengthOfData] = size(rawData(:,,:)); %returns dimensions,
    amount=#rows and length=#columns

    toc

    for count = 1:amountOfData %for loop: count will go through each
    row from 1 until the number of rows (as determined in line 2)

        ind = strfind(rawData(count,1:23), 'AGTCGTAAAC');

        ind1 = strfind(rawData(count,80:100), 'CGATTTTTCCT');

        if ~isempty(ind) && ~isempty(ind1)

            if (ind1+80-ind-11)== 60

```

```

        if ~isempty(strfind(rawData(count, ind+10:ind+14),
'AACCTC')) %A1
            barcodeA1 = 1;
        elseif ~isempty(strfind(rawData(count, ind+10:ind+14),
'AACCT')) %A2
            barcodeA1 = 2;
        elseif ~isempty(strfind(rawData(count, ind+10:ind+14),
'ATGTC')) %A3
            barcodeA1 = 3;
        elseif ~isempty(strfind(rawData(count, ind+10:ind+14),
'ATGCT')) %A4
            barcodeA1 = 4;
        elseif ~isempty(strfind(rawData(count, ind+10:ind+14),
'ACATC')) %A5
            barcodeA1 = 5;
        elseif ~isempty(strfind(rawData(count, ind+10:ind+14),
'ACACT')) %A6
            barcodeA1 = 6;
        elseif ~isempty(strfind(rawData(count, ind+10:ind+14),
'AGTTC')) %A7
            barcodeA1 = 7;
        elseif ~isempty(strfind(rawData(count, ind+10:ind+14),
'AGTCT')) %A8
            barcodeA1 = 8;
        elseif sum(nt2int(char(cellstr(rawData(count,
ind+10:ind+14))))==15)~=0 %nt2int converts nucleotides to a number and
'N' converts to 15 - this looks for N's
            barcodeA1 = 9;
        else
            barcodeA1 = 10;
        end

        if ~isempty(strfind(rawData(count, ind+15:ind+19),
'TCACAC')) %B1
            barcodeB1 = 1;
        elseif ~isempty(strfind(rawData(count, ind+15:ind+19),
'TCCAA')) %B2

```

```

        barcodeB1 = 2;

        elseif ~isempty(strfind(rawData(count, ind+15:ind+19),
'TCGTA')) %B3

            barcodeB1 = 3;

            elseif ~isempty(strfind(rawData(count, ind+15:ind+19),
'TCTGA')) %B4

                barcodeB1 = 4;

                elseif ~isempty(strfind(rawData(count, ind+15:ind+19),
'CTACA')) %B5

                    barcodeB1 = 5;

                    elseif ~isempty(strfind(rawData(count, ind+15:ind+19),
'CTCAA')) %B6

                        barcodeB1 = 6;

                        elseif ~isempty(strfind(rawData(count, ind+15:ind+19),
'CTGTA')) %B7

                            barcodeB1 = 7;

                            elseif ~isempty(strfind(rawData(count, ind+15:ind+19),
'CTTGA')) %B8

                                barcodeB1 = 8;

                                elseif sum(nt2int(char(cellstr(rawData(count,
ind+15:ind+19))))==15)~=0

                                    barcodeB1 = 9;

                                    else

                                        barcodeB1 = 10;

                                        end

                                if ~isempty(strfind(rawData(count, ind+25:ind+29),
'TCACA')) %B1

                                    barcodeB2 = 1;

                                    elseif ~isempty(strfind(rawData(count, ind+25:ind+29),
'TCCAA')) %B2

                                        barcodeB2 = 2;

                                        elseif ~isempty(strfind(rawData(count, ind+25:ind+29),
'TCGTA')) %B3

                                            barcodeB2 = 3;

                                            elseif ~isempty(strfind(rawData(count, ind+25:ind+29),

```

```

'TCTGA')) %B4
        barcodeB2 = 4;
    elseif ~isempty(strfind(rawData(count, ind+25:ind+29),
'CTACA')) %B5
        barcodeB2 = 5;
    elseif ~isempty(strfind(rawData(count, ind+25:ind+29),
'CTCAA')) %B6
        barcodeB2 = 6;
    elseif ~isempty(strfind(rawData(count, ind+25:ind+29),
'CTGTA')) %B7
        barcodeB2 = 7;
    elseif ~isempty(strfind(rawData(count, ind+25:ind+29),
'CTTGA')) %B8
        barcodeB2 = 8;
    elseif sum(nt2int(char(cellstr(rawData(count,
ind+25:ind+29))))==15)~=0
        barcodeB2 = 9;
    else
        barcodeB2 = 10;
    end

    if ~isempty(strfind(rawData(count, ind+35:ind+39),
'TCACA')) %B1
        barcodeB3 = 1;
    elseif ~isempty(strfind(rawData(count, ind+35:ind+39),
'TCCAA')) %B2
        barcodeB3 = 2;
    elseif ~isempty(strfind(rawData(count, ind+35:ind+39),
'TCGTA')) %B3
        barcodeB3 = 3;
    elseif ~isempty(strfind(rawData(count, ind+35:ind+39),
'TCTGA')) %B4
        barcodeB3 = 4;
    elseif ~isempty(strfind(rawData(count, ind+35:ind+39),
'CTACA')) %B5
        barcodeB3 = 5;

```

```

elseif ~isempty(strfind(rawData(count, ind+35:ind+39),
'CTCAA')) %B6
    barcodeB3 = 6;
elseif ~isempty(strfind(rawData(count, ind+35:ind+39),
'CTGTA')) %B7
    barcodeB3 = 7;
elseif ~isempty(strfind(rawData(count, ind+35:ind+39),
'CTTGA')) %B8
    barcodeB3 = 8;
elseif sum(nt2int(char(cellstr(rawData(count,
ind+35:ind+39))))==15)~=0
    barcodeB3 = 9;
else
    barcodeB3 = 10;
end

if ~isempty(strfind(rawData(count, ind+45:ind+49),
'TCACA')) %B1
    barcodeB4 = 1;
elseif ~isempty(strfind(rawData(count, ind+45:ind+49),
'TCCAA')) %B2
    barcodeB4 = 2;
elseif ~isempty(strfind(rawData(count, ind+45:ind+49),
'TCGTA')) %B3
    barcodeB4 = 3;
elseif ~isempty(strfind(rawData(count, ind+45:ind+49),
'TCTGA')) %B4
    barcodeB4 = 4;
elseif ~isempty(strfind(rawData(count, ind+45:ind+49),
'CTACA')) %B5
    barcodeB4 = 5;
elseif ~isempty(strfind(rawData(count, ind+45:ind+49),
'CTCAA')) %B6
    barcodeB4 = 6;
elseif ~isempty(strfind(rawData(count, ind+45:ind+49),
'CTGTA')) %B7

```

```

        barcodeB4 = 7;

        elseif ~isempty(strfind(rawData(count, ind+45:ind+49),
'CTTGA')) %B8

            barcodeB4 = 8;

            elseif sum(nt2int(char(cellstr(rawData(count,
ind+45:ind+49))))==15)~=0

                barcodeB4 = 9;

            else

                barcodeB4 = 10;

            end

        if ~isempty(strfind(rawData(count, ind+55:ind+59),
'TCACA')) %B1

            barcodeB5 = 1;

            elseif ~isempty(strfind(rawData(count, ind+55:ind+59),
'TCCAA')) %B2

                barcodeB5 = 2;

            elseif ~isempty(strfind(rawData(count, ind+55:ind+59),
'TCGTA')) %B3

                barcodeB5 = 3;

            elseif ~isempty(strfind(rawData(count, ind+55:ind+59),
'TCTGA')) %B4

                barcodeB5 = 4;

            elseif ~isempty(strfind(rawData(count, ind+55:ind+59),
'CTACA')) %B5

                barcodeB5 = 5;

            elseif ~isempty(strfind(rawData(count, ind+55:ind+59),
'CTCAA')) %B6

                barcodeB5 = 6;

            elseif ~isempty(strfind(rawData(count, ind+55:ind+59),
'CTGTA')) %B7

                barcodeB5 = 7;

            elseif ~isempty(strfind(rawData(count, ind+55:ind+59),
'CTTGA')) %B8

                barcodeB5 = 8;

            elseif sum(nt2int(char(cellstr(rawData(count,

```



```

ind+55:ind+59)))==15)~=0

        barcodeB5 = 9;

    else

        barcodeB5 = 10;

    end

    if ~isempty(strfind(rawData(count, ind+65:ind+69),
'TCACA')) %B1

        barcodeB6 = 1;

    elseif ~isempty(strfind(rawData(count, ind+65:ind+69),
'TCCAA')) %B2

        barcodeB6 = 2;

    elseif ~isempty(strfind(rawData(count, ind+65:ind+69),
'TCGTA')) %B3

        barcodeB6 = 3;

    elseif ~isempty(strfind(rawData(count, ind+65:ind+69),
'TCTGA')) %B4

        barcodeB6 = 4;

    elseif ~isempty(strfind(rawData(count, ind+65:ind+69),
'CTACA')) %B5

        barcodeB6 = 5;

    elseif ~isempty(strfind(rawData(count, ind+65:ind+69),
'CTCAA')) %B6

        barcodeB6 = 6;

    elseif ~isempty(strfind(rawData(count, ind+65:ind+69),
'CTGTA')) %B7

        barcodeB6 = 7;

    elseif ~isempty(strfind(rawData(count, ind+65:ind+69),
'CTTGA')) %B8

        barcodeB6 = 8;

    elseif sum(nt2int(char(cellstr(rawData(count,
ind+65:ind+69))))==15)~=0

        barcodeB6 = 9;

    else

```

```

        barcodeB6 = 10;
    end

    if ~isempty(strfind(rawData(count, ind+20:ind+24),
'AACTC')) %A1
        barcodeA2 = 1;
    elseif ~isempty(strfind(rawData(count, ind+20:ind+24),
'AACTT')) %A2
        barcodeA2 = 2;
    elseif ~isempty(strfind(rawData(count, ind+20:ind+24),
'ATGTC')) %A3
        barcodeA2 = 3;
    elseif ~isempty(strfind(rawData(count, ind+20:ind+24),
'ATGCT')) %A4
        barcodeA2 = 4;
    elseif ~isempty(strfind(rawData(count, ind+20:ind+24),
'ACATC')) %A5
        barcodeA2 = 5;
    elseif ~isempty(strfind(rawData(count, ind+20:ind+24),
'ACACT')) %A6
        barcodeA2 = 6;
    elseif ~isempty(strfind(rawData(count, ind+20:ind+24),
'AGTTC')) %A7
        barcodeA2 = 7;
    elseif ~isempty(strfind(rawData(count, ind+20:ind+24),
'AGTCT')) %A8
        barcodeA2 = 8;
    elseif sum(nt2int(char(cellstr(rawData(count,
ind+20:ind+24))))==15)~=0
        barcodeA2 = 9;
    else
        barcodeA2 = 10;
    end

    if ~isempty(strfind(rawData(count, ind+30:ind+34),
'AACTC')) %A1

```

```

        barcodeA3 = 1;

        elseif ~isempty(strfind(rawData(count, ind+30:ind+34),
'AACCT')) %A2
            barcodeA3 = 2;

            elseif ~isempty(strfind(rawData(count, ind+30:ind+34),
'ATGTC')) %A3
                barcodeA3 = 3;

                elseif ~isempty(strfind(rawData(count, ind+30:ind+34),
'ATGCT')) %A4
                    barcodeA3 = 4;

                    elseif ~isempty(strfind(rawData(count, ind+30:ind+34),
'ACATC')) %A5
                        barcodeA3 = 5;

                        elseif ~isempty(strfind(rawData(count, ind+30:ind+34),
'ACACT')) %A6
                            barcodeA3 = 6;

                            elseif ~isempty(strfind(rawData(count, ind+30:ind+34),
'AGTTC')) %A7
                                barcodeA3 = 7;

                                elseif ~isempty(strfind(rawData(count, ind+30:ind+34),
'AGTCT')) %A8
                                    barcodeA3 = 8;

                                    elseif sum(nt2int(char(cellstr(rawData(count,
ind+30:ind+34))))==15)~=0
                                        barcodeA3 = 9;

                                        else
                                            barcodeA3 = 10;

                                        end

'AACCTC')) %A1
        if ~isempty(strfind(rawData(count, ind+40:ind+44),
'AACCT')) %A2
            barcodeA4 = 1;

            elseif ~isempty(strfind(rawData(count, ind+40:ind+44),
'AACCT')) %A2
                barcodeA4 = 2;

                elseif ~isempty(strfind(rawData(count, ind+40:ind+44),

```

```

'ATGTC')) %A3
        barcodeA4 = 3;
    elseif ~isempty(strfind(rawData(count, ind+40:ind+44),
'ATGCT')) %A4
        barcodeA4 = 4;
    elseif ~isempty(strfind(rawData(count, ind+40:ind+44),
'ACATC')) %A5
        barcodeA4 = 5;
    elseif ~isempty(strfind(rawData(count, ind+40:ind+44),
'ACACT')) %A6
        barcodeA4 = 6;
    elseif ~isempty(strfind(rawData(count, ind+40:ind+44),
'AGTTC')) %A7
        barcodeA4 = 7;
    elseif ~isempty(strfind(rawData(count, ind+40:ind+44),
'AGTCT')) %A8
        barcodeA4 = 8;
    elseif sum(nt2int(char(cellstr(rawData(count,
ind+40:ind+44))))==15)~=0
        barcodeA4 = 9;
    else
        barcodeA4 = 10;
    end

    if ~isempty(strfind(rawData(count, ind+50:ind+54),
'AACTC')) %A1
        barcodeA5 = 1;
    elseif ~isempty(strfind(rawData(count, ind+50:ind+54),
'AACCT')) %A2
        barcodeA5 = 2;
    elseif ~isempty(strfind(rawData(count, ind+50:ind+54),
'ATGTC')) %A3
        barcodeA5 = 3;
    elseif ~isempty(strfind(rawData(count, ind+50:ind+54),
'ATGCT')) %A4
        barcodeA5 = 4;

```

```

elseif ~isempty(strfind(rawData(count, ind+50:ind+54),
'ACATC')) %A5
    barcodeA5 = 5;
elseif ~isempty(strfind(rawData(count, ind+50:ind+54),
'ACACT')) %A6
    barcodeA5 = 6;
elseif ~isempty(strfind(rawData(count, ind+50:ind+54),
'AGTTC')) %A7
    barcodeA5 = 7;
elseif ~isempty(strfind(rawData(count, ind+50:ind+54),
'AGTCT')) %A8
    barcodeA5 = 8;
elseif sum(nt2int(char(cellstr(rawData(count,
ind+50:ind+54))))==15)~=0
    barcodeA5 = 9;
else
    barcodeA5 = 10;
end

if ~isempty(strfind(rawData(count, ind+60:ind+64),
'AATC')) %A1
    barcodeA6 = 1;
elseif ~isempty(strfind(rawData(count, ind+60:ind+64),
'AACCT')) %A2
    barcodeA6 = 2;
elseif ~isempty(strfind(rawData(count, ind+60:ind+64),
'ATGTC')) %A3
    barcodeA6 = 3;
elseif ~isempty(strfind(rawData(count, ind+60:ind+64),
'ATGCT')) %A4
    barcodeA6 = 4;
elseif ~isempty(strfind(rawData(count, ind+60:ind+64),
'ACATC')) %A5
    barcodeA6 = 5;
elseif ~isempty(strfind(rawData(count, ind+60:ind+64),
'ACACT')) %A6

```

```

        barcodeA6 = 6;

        elseif ~isempty(strfind(rawData(count, ind+60:ind+64),
'AGTTC')) %A7

            barcodeA6 = 7;

            elseif ~isempty(strfind(rawData(count, ind+60:ind+64),
'AGTCT')) %A8

                barcodeA6 = 8;

                elseif sum(nt2int(char(cellstr(rawData(count,
ind+60:ind+64))))==15)~=0

                    barcodeA6 = 9;

                else

                    barcodeA6 = 10;

                end

        tableOfResults(barcodeA1,1) = tableOfResults(barcodeA1,1) + 1;
        tableOfResults(barcodeB1,2) = tableOfResults(barcodeB1,2) + 1;
        tableOfResults(barcodeA2,3) = tableOfResults(barcodeA2,3) + 1;

        tableOfResults(barcodeB2,4) =
        tableOfResults(barcodeB2,4) + 1;
        tableOfResults(barcodeA3,5) = OfResults(barcodeA3,5) +
        1;

        tableOfResults(barcodeB3,6) = tableOfResults(barcodeB3,6) + 1;
        tableOfResults(barcodeA4,7) = tableOfResults(barcodeA4,7) + 1;
        tableOfResults(barcodeB4,8) = tableOfResults(barcodeB4,8) + 1;
        tableOfResults(barcodeA5,9) = tableOfResults(barcodeA5,9) + 1;
        tableOfResults(barcodeB5,10) = tableOfResults(barcodeB5,10) + 1;
        tableOfResults(barcodeA6,11) = tableOfResults(barcodeA6,11) + 1;
        tableOfResults(barcodeB6,12) = tableOfResults(barcodeB6,12) + 1;

        if (barcodeA1 < 9 && barcodeA2 <9 && barcodeA3 <9 && barcodeA4
<9 && barcodeA5 <9 && barcodeA6 < 9 && barcodeB1 <9 && barcodeB2 <9 &&
barcodeB3 <9 && barcodeB4 <9 && barcodeB5 <9 && barcodeB6 <9)

            countTotal = countTotal + 1;

```

```

        fprintf(fid, '%s \n', rawData(count,ind+10:ind+69)); %prints
(to file) data with codes that were correctly assigned for all positions

        elseif (barcodeA1 == 9 || barcodeA2 == 9 || barcodeA3 == 9 ||
barcodeA4 ==9 || barcodeA5 == 9 || barcodeA6 == 9 || barcodeB1 == 9 ||
barcodeB2 == 9 || barcodeB3 == 9 || barcodeB4 == 9 || barcodeB5 ==9 ||
barcodeB6 == 9)

            HasN = HasN + 1;

        else

            invalidData = invalidData + 1;

            %fprintf(fid, '%s \n', rawData(count,:));

        end

        else

            InDel = InDel + 1;

        end

        else

            PrimMut = PrimMut + 1;

        end

    end

end

save(name, 'tableOfResults', '-ascii', '-tabs')

end

```

MATLAB Program for Analyzing Library Diversity from a High-Throughput

Sequencing Result

```

clear uReads

clear dupReads

perfectReads = importdata('Jia-PerfectLibraryMembers_2.txt');

[uReads,~,n] = unique(perfectReads);

numUnique = numel(uReads)

readFreq = accumarray(n(:),1);

```

```

figure(); hist(readFreq, unique(readFreq));

xlabel('Occurrences', 'FontSize',24); ylabel('Number of Sequence
Reads','FontSize', 24);

title('Read Occurrences','FontSize',20)

d = readFreq > 1;
dupReads = uReads(d) '
dupFreq = readFreq(d) '

numDupReads = numel(dupReads)

```

MATLAB Program for Searching Unexpected Biotinylation Codon from a High-Throughput Sequencing Result

```

tic

countTotal = 0;
countBio = 0;
countLib = 0;
PrimMut = 0;
posCtrl = 0;
falsePos = 0;

fid1 = fopen('Jia-FalsePositiveMembers_2.txt', 'at');
fid2 = fopen('Jia-PositiveControlMembers_2.txt', 'at');

for x = 1:200
    z = x*200000
    y = z-200000+1
    toc
    [Header, Sequence] = fastqread('2_S2_L001_R1_001.fastq', 'Blockread', [y

```



```

z]);

rawData = char(Sequence); %converts cell array
to character array, putting each letter in its own column

clear Sequence

clear Header

[amountOfData lengthOfData] = size(rawData(:, :)); %returns dimensions,
amount=#rows and length=#columns

toc

for count = 1:amountOfData %for loop: count will go through each
row from 1 until the number of rows (as determined in line 2)

    ind = strfind(rawData(count, 1:23), 'AGTCGTAAAC');

    if ~isempty(ind)

        if ~isempty(strfind(rawData(count, ind+10:ind+79),
'GCATT')) %Biotin codon

            countBio = countBio + 1;

            if ~isempty(strfind(rawData(count, ind+70:ind+77),
'GCATTAAT')) %Postive cotrol count

                posCtrl = posCtrl + 1;

                fprintf(fid2, '%s \n',
rawData(count, ind+10:ind+77)); %Print postive control sequences

            else

                falsePos = falsePos + 1; %false positive count

                fprintf(fid1, '%s \n',
rawData(count, ind+10:ind+78)); %Print false positive sequences

            end

        else

            countLib = countLib + 1;

        end
    end
end

```

```
        else
            PrimMut = PrimMut + 1;
        end
    end
end

end
```

References

1. Brudno, Y., Birnbaum, M. E., Kleiner, R. E. & Liu, D. R. An in vitro translation, selection and amplification system for peptide nucleic acids. *Nat. Chem. Biol.* **6**, 148–55 (2010).
2. Lutz, J.-F., Ouchi, M., Liu, D. R. & Sawamoto, M. Sequence-controlled polymers. *Science* **341**, 628–636 (2013).
3. Seebach, D., Beck, A. K. & Bierbaum, D. J. The World of b- and g-Peptides Comprised of Homologated Proteinogenic Amino Acids and Other Components. *Chem. Biodiv.* **1**, 1111–1239 (2004).
4. Petersson, E. J. & Schepartz, A. Toward beta-amino acid proteins: design, synthesis, and characterization of a fifteen kilodalton beta-peptide tetramer. *J. Am. Chem. Soc.* **130**, 821–3 (2008).
5. Goodman, C. M., Choi, S., Shandler, S. & DeGrado, W. F. Foldamers as versatile frameworks for the design and evolution of function. *Nat. Chem. Biol.* **3**, 252–62 (2007).
6. Martinek, T. a. & Fulop, F. Side-chain control of beta-peptide secondary structures. Design principles. *Eur. J. Biochem.* **270**, 3657–3666 (2003).
7. Zuckermann, R. N., Kerr, J. M., Kent, S. B. H. & Moost, W. H. Efficient Method for the Preparation of Peptoids [Oligo(N-substituted glycines)] by Submonomer Solid-Phase Synthesis. *J. Am. Chem. Soc.* **114**, 10646–10647 (1992).
8. Horne, W. S. & Gellman, S. H. Foldamers with heterogeneous backbones. *Acc. Chem. Res.* **41**, 1399–408 (2008).
9. Bautista, A. D., Craig, C. J., Harker, E. a & Schepartz, A. Sophistication of foldamer form and function in vitro and in vivo. *Curr. Opin. Chem. Biol.* **11**, 685–92 (2007).
10. Pilsl, L. K. a & Reiser, O. α/β -Peptide foldamers: state of the art. *Amino Acids* **41**, 709–18 (2011).
11. Horne, W. S., Price, J. L., Keck, J. L. & Gellman, S. H. Helix bundle quaternary structure from alpha/beta-peptide foldamers. *J. Am. Chem. Soc.* **129**, 4178–80 (2007).
12. Goodman, J. L., Molski, M. a, Qiu, J. & Schepartz, A. Tetrameric beta(3)-peptide bundles. *Chembiochem* **9**, 1576–8 (2008).

13. Larbanoix, L. *et al.* Design and evaluation of a 6-mer amyloid-beta protein derived phage display library for molecular targeting of amyloid plaques in Alzheimer's disease: Comparison with two cyclic heptapeptides derived from a randomized phage display library. *Peptides* **32**, 1232–43 (2011).
14. Qiu, J. X., Petersson, E. J., Matthews, E. E. & Schepartz, A. Toward beta-amino acid proteins: a cooperatively folded beta-peptide quaternary structure. *J. Am. Chem. Soc.* **128**, 11338–9 (2006).
15. Kritzer, J. a *et al.* Relationship between side chain structure and 14-helix stability of beta3-peptides in water. *J. Am. Chem. Soc.* **127**, 167–78 (2005).
16. Raguse, T. L., Lai, J. R. & Gellman, S. H. Environment-independent 14-helix formation in short beta-peptides: striking a balance between shape control and functional diversity. *J. Am. Chem. Soc.* **125**, 5592–3 (2003).
17. Daura, X. *et al.* The beta-peptide hairpin in solution: conformational study of a beta-hexapeptide in methanol by NMR spectroscopy and MD simulation. *J. Am. Chem. Soc.* **123**, 2393–404 (2001).
18. Murray, J. K. & Gellman, S. H. Microwave-assisted parallel synthesis of a 14-helical beta-peptide library. *J. Comb. Chem.* **8**, 58–65 (2006).
19. Seelig, B. mRNA display for the selection and evolution of enzymes from in vitro-translated protein libraries. *Nat. Protoc.* **6**, 540–52 (2011).
20. Wright, M. C. & Joyce, G. F. Continuous in Vitro Evolution of Catalytic Function. *Science* **276**, 614–617 (1997).
21. Joyce, G. F. Nonenzymatic Template-directed Synthesis of Informational Macromolecules. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 41–51 (1987).
22. Daguer, J. P. *et al.* DNA-templated combinatorial assembly of small molecule fragments amenable to selection/amplification cycles. *Chem. Sci.* **2**, 625 (2011).
23. Chan, T. R., Hilgraf, R., Sharpless, K. B. & Fokin, V. V. Polytriazoles as copper(I)-stabilizing ligands in catalysis. *Org. Lett.* **6**, 2853–5 (2004).

Chapter Four

DNA Ligase-Mediated Translation of DNA into Highly Functionalized Nucleic Acid Synthetic Polymers

Ryan Hili, Jia Niu, and David R. Liu

I collaborated with Ryan Hili in developing the T4 DNA Ligase-mediated translation system. I performed the polymerization experiments with doubly modified trinucleotides. Ryan Hili led the development of the DNA-ligase mediated polymerization and the integration of this translation strategy into an *in vitro* selection system.

Adapted from: Hili, R., Niu, J., and Liu, D. R. DNA Ligase-Mediated Translation of DNA Into Densely Functionalized Nucleic Acid Polymers. *J. Am. Chem. Soc.* 135, 98-101 (2013).

4.1 Introduction

In the previous chapters, we described an enzyme-free translation system to generate non-nucleic acid polymers using DNA as templates, and the integration of this translation strategy into iterative cycles of *in vitro* selections for discovery of novel functional synthetic polymers. The enzyme-free approach has established the foundation of exploring unprecedented functional and structural space, but it is also limited by the low translation efficiency compared to enzymatic systems, and complex substrate synthesis that are difficult to scale up. We reason that by expanding the chemical scope of side chain modifications, a modified nucleic acid polymer could potentially approach the breadth and depth of the functional space that proteins have achieved. Such densely modified nucleic acid polymers would have the advantages of both the structural diversity of side chain functional groups and the versatility and high efficiency of the enzymatic systems to generate nucleic acid polymers.

All previously reported strategies for sequence-specific incorporation of non-natural functionality throughout a nucleic acid polymer rely on DNA or RNA polymerases to incorporate modified nucleotides.¹⁻¹² While effective in generating modified nucleic acid polymers, these systems limit the number of different side-chain modifications to maximally four in single-stranded DNA (with the exception of using an expanded genetic code, in which the potential modifications could reach up to six¹³), and eight in double-stranded DNA.¹⁴ In addition, the nature of these modifications is also limited by the tolerance of the polymerase active site. A new strategy is needed to be able to incorporate more different functional side-chain modifications in a sequence-defined fashion into nucleic acid polymers.

4.2 Design of DNA-Templated Polymerization of Short Modified Oligonucleotides

We hypothesized that a ligase-mediated DNA templated polymerization of short 5'-phosphorylated oligonucleotides containing various modifications might enable access to these desired functionalized polymers. Our attention was directed toward T4 DNA ligase as a potential enzyme suitable for the translation of DNA into modified DNA. As it continues to serve as one of the most widely used enzymes in molecular biology, it has been the subject of several in-depth structural, functional, and mechanistic studies.¹⁵⁻²¹ Its high stability and sequence specificity together with its ability to perform consecutive ligations of 5'-phosphorylated hexanucleotides along a DNA template has been harnessed in several sequencing technologies.^{22,23} While the tolerance of T4 DNA ligase for accepting modified DNA has not been extensively explored, modifications at the ligation site have been reported for both the ligation strands and the template strand.^{24,25}

Inspired by the trinucleotide codon system used in Nature during the translation of mRNA into proteins on the ribosome, we explored the possibility of using 5'-phosphorylated trinucleotides as our functionalized monomers to be polymerized using T4 DNA ligase. Such a system would enable up to 64 different modifications to be incorporated throughout a nucleic acid polymer, or the possibility of codon redundancy, if desired. While the shortest oligonucleotide length reported to undergo ligation using T4 DNA ligase was a pentamer,²⁶ we hypothesized that the optimizations of the ligation reaction conditions might enable the polymerization of modified trinucleotides.

We envisioned the translation process occurring within a reading frame defined by a set of primers, namely a 5'-phosphorylated initiation primer and a non-phosphorylated termination primer (Figure 4.1a). The role of the initiation primer is to specify the beginning of the reading frame, while the role of the termination primer is to specify the end of the reading frame and

preclude the formation of blunt- or cohesive-end ligation byproducts. Between the set of primers lies the specified reading frame: the set of codons to be translated by T4 DNA ligase into a functionalized nucleic acid polymer. 5'-Phosphorylated trinucleotides were synthesized using standard automated oligonucleotide synthesis with commercially available amine-modified nucleoside phosphoramidites and 5'-phosphorylation reagents. The amine group of the nucleobase can potentially serve as a handle to incorporate various functional groups using well-established bioconjugation chemistries such as amine addition to cyclic anhydrides, isothiocyanides, and activated esters (Figure 4.1b).

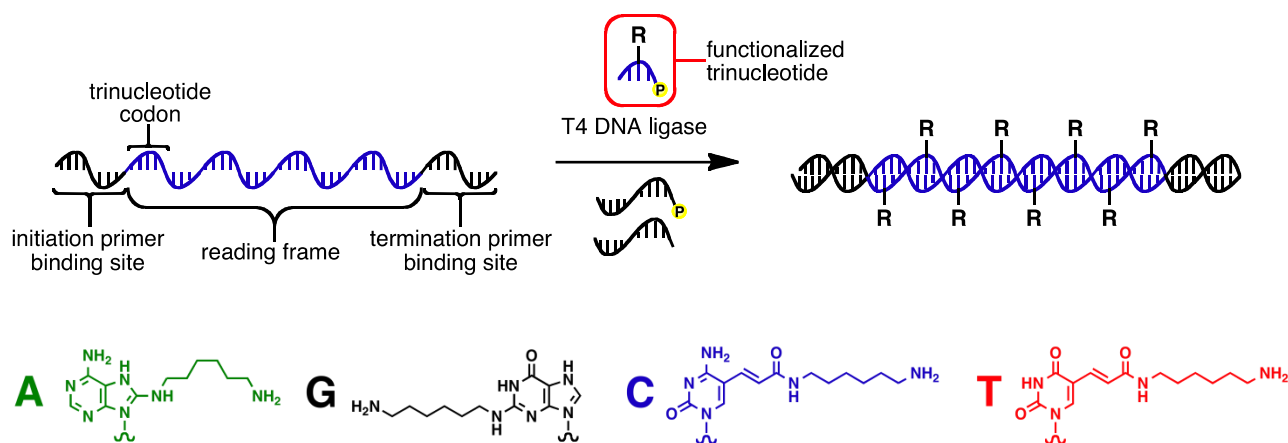


Figure 4.1. (a) General strategy for the translation of DNA into functionalized DNA using T4 DNA ligase-mediated sequence-defined polymerization of functionalized trinucleotides. (b) General structures of modified bases of trinucleotides.

4.3 Optimization of the Reaction Conditions for the Ligase-Mediated Polymerization of Nucleobase-Modified Trinucleotides

In order to determine the optimal position of the side-chain modifications, we synthesized a series of trinucleotides with single or double modifications. The singly modified nucleobases all have amine modifications at either the first (5'-end), second, or third (3'-end)

position. The doubly modified trinucleotides have an amine- and an alkyne-modified base at two of the three nucleotides (Figure 4.2b). For all doubly modified trinucleotides, the T4 DNA ligase-mediated polymerizations were inefficient. We did not see any polymerizations for the 2,3- and 1,2-modified trinucleotides, probably due to the requirement of maintaining good contact of polymerase with the nucleobase at that position the nucleobase in catalysis. Albeit with compromised efficiency and kinetics, the 1,3-modified trinucleotide exhibited the ability to serve as the substrate of T4 DNA ligase, and the polymerization could reach ~80% yield of full-length over 40 hours under 20 °C (Figure 4.2c and d). Due to the lack of efficiency during the polymerizations of doubly modified trinucleotides, we decided to focus on using mono-modified trinucleotides in our further studies. We found that T4 DNA ligase had a strong preference for substitutions on the first nucleobase, with no observable polymerization occurring when modifications were at the second position (Figure 4.2a). When modifications were presented on the Watson-Crick face at any position, T4 DNA ligase failed to effect polymerization to any observable extent (Figure 4.2a).

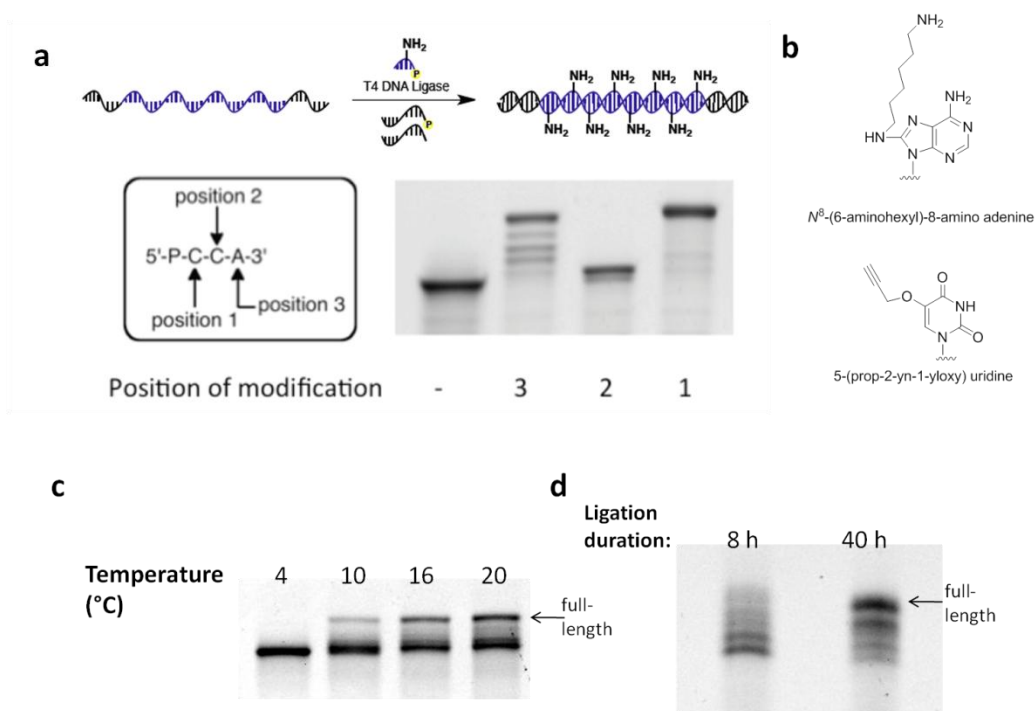


Figure 4.2. (a) Positional scan of the efficiency of ligase-mediated polymerization of singly modified trinucleotide substrates. (b) Structure of two modified bases for the doubly modified trinucleotides. (c) Polymerization of 1,3-modified trinucleotides under various temperatures. The polymerization duration was 16 hours. (d) Polymerization of 1,3-modified trinucleotides with extended reaction times.

During our initial optimization of the polymerization reaction, we observed that the molecular crowding reagent PEG 6000²⁷ was an essential component of the re-action to ensure complete conversion to full-length product. Furthermore, due to the required extended ligation times used during the polymerization, we observed that the addition of 0.1 mg/mL of BSA greatly improved the activity of T4 DNA ligase for reactions beyond 2 h.²⁷ Despite the melting temperature of the trinucleotides used during these experiments being <10 °C, the polymerization of the modified trinucleotides was found to be efficient up to 30 °C, highlighting the efficiency of T4 DNA ligase at ligating the modified trinucleotides under the optimized conditions. With the optimized conditions in hand, we chose to scale up the polymerization

process and analyze the translated single-stranded nucleic acid polymer by ESI-LC-MS, which was found to be consistent with the expected mass (calculated mass: 15771.1 Da, observed mass: 15770.6 Da).

We next challenged the optimized polymerization system with various modified trinucleotides of different nucleic acid sequences (Figure 4.3a). The modifications included functionality ranging from hydrogen bond donors/acceptors, Brønsted acids/bases, metal chelators, and hydrophobic groups. Each modified trinucleotide was assessed by T4 DNA ligase-mediated polymerization along a homo-octameric template and then analyzed by native PAGE (Figure 4.3b). All modified trinucleotides were polymerized efficiently, yielding products of the expected molecular weights.

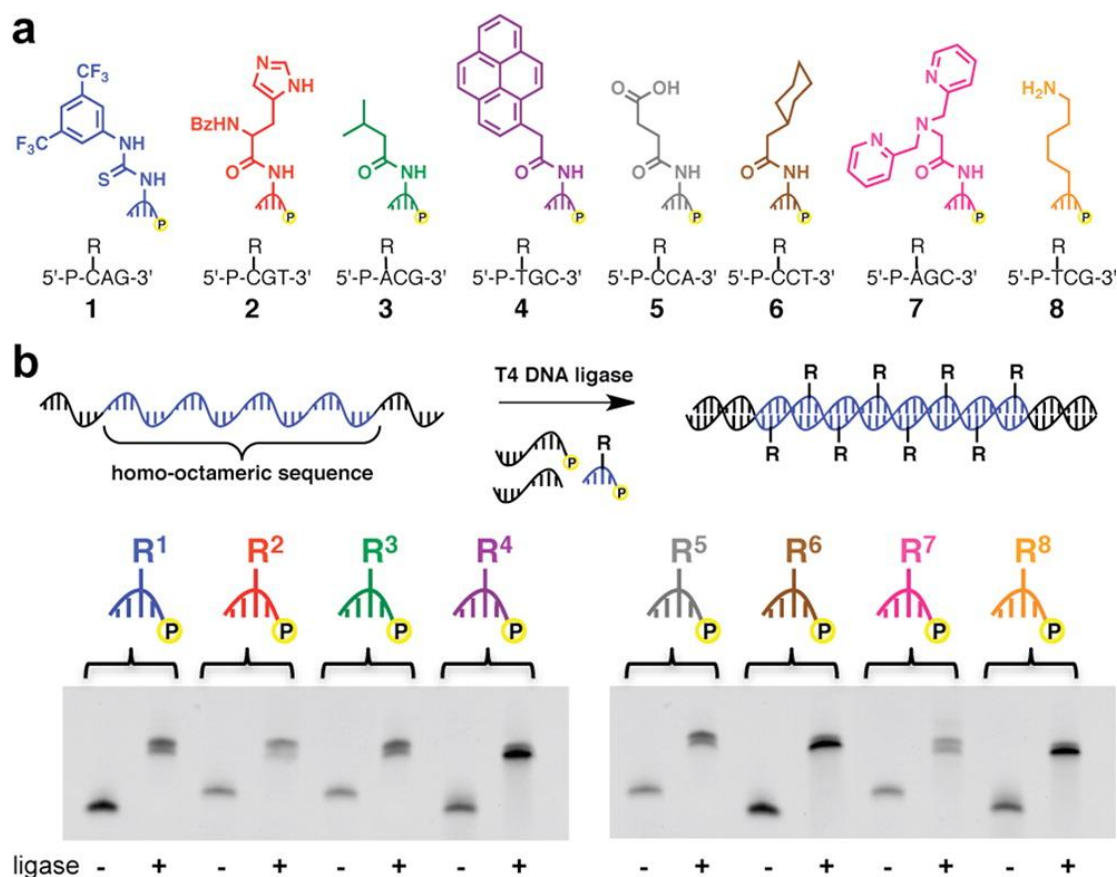


Figure 4.3. (a) Structure and sequence of functionalized trinucleotides. (b) Nondenaturing PAGE analysis of products from polymerization of the functionalized trinucleotides in (a) along their corresponding homo-octameric codon templates.

4.4. Sequence-Specificity of Ligase-Mediated Polymerization

To determine the sequence specificity of polymerization, we employed a chain-termination strategy. Since T4 DNA ligase requires 5'-phosphates to continue the templated polymerization of trinucleotides, incorporating trinucleotides that lack a 5'-phosphate into the reaction should terminate the polymerization, resulting in truncation products. If T4 DNA ligase incorporates the trinucleotides in a sequence-specific manner, then chain-termination should

occur only at the codon specifying the non-phosphorylated trinucleotide. If sequence-specificity is low, then polymerization should generate nucleic acid polymers of undesired length. This could arise either by read-through of the terminator codon, or by non-specific incorporation of the terminator trinucleotide at non-terminating codons. Our specificity assay involved the use of a hetero-octameric template comprising all eight codons flanked by 12-nt long initiation and termination primers. Seven 5'-phosphorylated functionalized trinucleotides and one functionalized trinucleotide lacking a 5'-phosphate were polymerized along the template, and the reaction products were analyzed by native PAGE. Only one termination product was observed at the anticipated mass for each chain-termination reaction, signifying high sequence specificity for the incorporation of each modified trinucleotide (Figure 4.4).

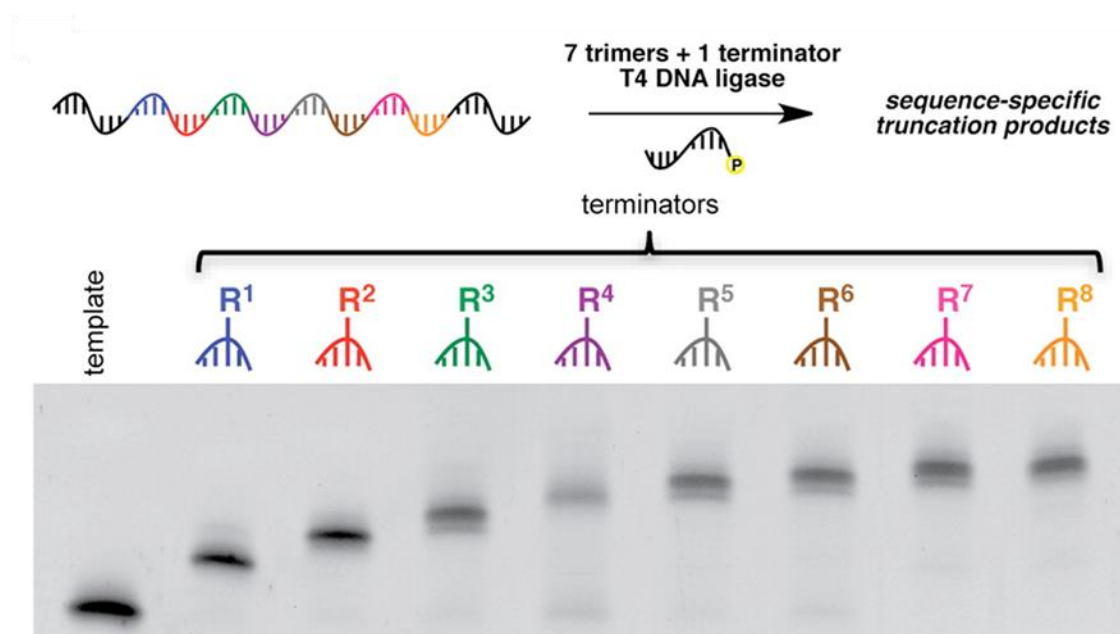


Figure 4.4. Nondenaturing PAGE analysis of sequence specificity using a chain-termination method.

4.5. Polymerization to Extended Lengths

We next sought to assess the efficiency of polymerization along templates of increasing sequence length. Templates comprising reading frames ranging from 10-50 codons for the isopropyl-modified trinucleotide flanked by 12 nt primer sets were synthesized. The polymerizations were performed at 16 °C for 12 h under the condition described above. Full-length products were generated in high yields for up to 50 codons on a homomeric template, underscoring the highly efficient polymerization ability of T4 DNA ligase.

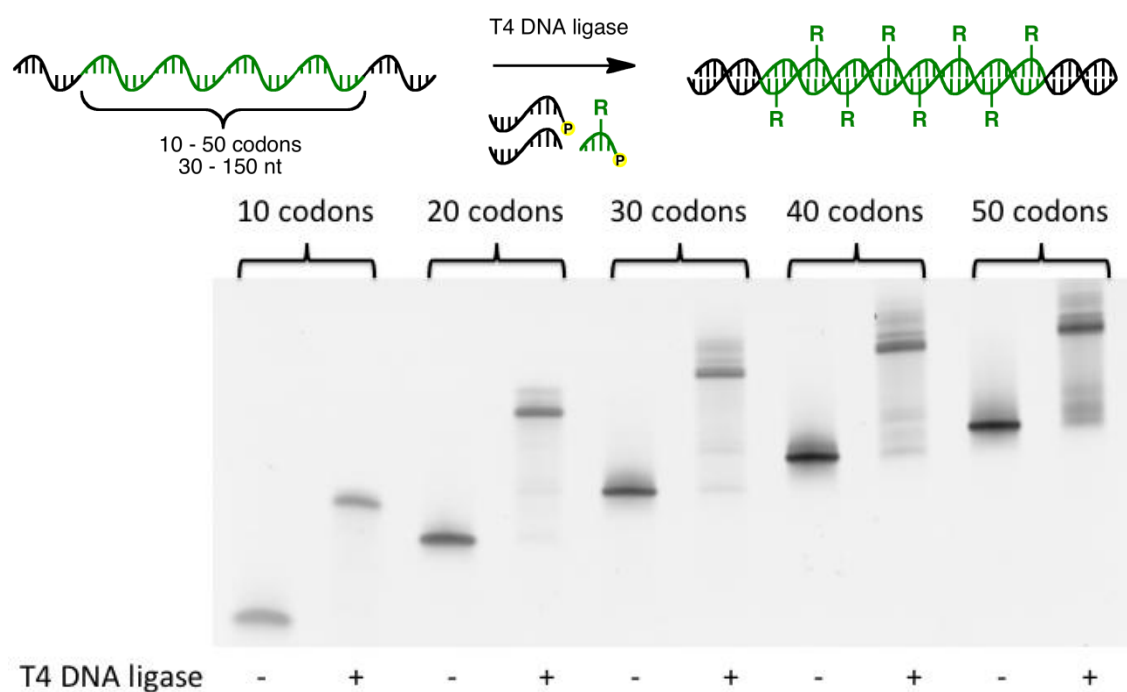


Figure 4.5. Native PAGE analysis of polymerization process along templates of increasing sequence length

4.6. Reverse Translation Using DNA Polymerase-Mediated Primer Extension

In order for this method to be applied toward the discovery of functional nucleic acids through in vitro selection, the modified nucleic acid polymer must be amenable to enzymatic

reverse-translation back into canonical DNA. Recent efforts have demonstrated that Family B DNA polymerases are particularly suited for primer extension along modified nucleic acid templates. Encouraged by these findings, we sought to examine a panel of polymerases for their efficacy at reverse transcribing our modified ssDNA back into their original unmodified DNA template form. The modified ssDNA that we chose to study was a hetero-octameric polymer of modified trinucleotides generated by T4 DNA ligase-mediated polymerization, representing each functional group once, and flanked by 18 nt primers. Consistent with previous findings,^{28,29} we observed that Deep Vent (exo-) polymerase performed primer extension with excellent conversions after a 30 minute incubation at 70 °C (Figure 4.6). While KOD XL (exo-) generated a small amount of detectable product, Therminator, pfu (exo-), Taq, and Tth polymerases failed to generate any desired product. To determine the sequence specificity of the primer extension, we PCR amplified the PAGE purified product of the Deep Vent (exo-) primer extension and performed Sanger sequencing. The sequence of the template that was regenerated from the primer extension process using Deep Vent (exo-) was consistent with the initial template sequence, demonstrating that both the translation and reverse transcription processes occurred with high sequence specificity.

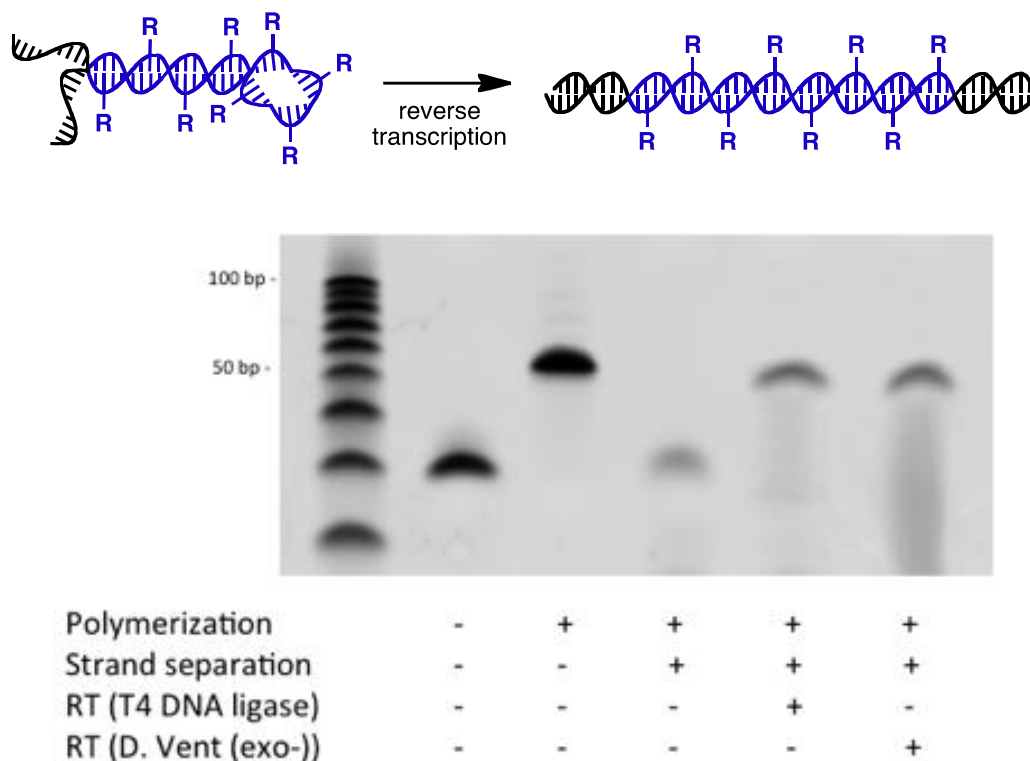


Figure 4.6. T4 DNA ligase-mediated translation followed by reverse translation back to native DNA, a primer extension reaction catalyzed by Deep Vent DNA polymerase.

4.7 Iterative Cycles of Translation, Selection, Template Re-generation, and Re-translation:

A Mock Selection.

To test a full cycle of translation, selection, template re-generation, and amplification in a library format, we sought to select a modified DNA containing a pharmacophore that binds carbonic anhydrase II from a translated library of highly functionalized DNA (Figure 4.7a). We generated a library of DNA templates using split-and-pool synthesis, which encoded seven trinucleotide substrates (1–7, Figure 4.3a) across eight coding positions. The theoretical complexity of this library is 5.8×10^6 . Next, we performed a translation and mock selection

using a solution containing this template library and $1/5.8 \times 10^6$ th of one equivalent of a positive control template that uniquely encodes the carbonic anhydrase II inhibitor Gly-Leu-4-carboxybenzene-sulfonamide³⁰ attached to substrate 8 and that contains an MluI restriction site to monitor the enrichment of the sequence by digestion and PAGE analysis. Following four iterated rounds of translation, strand separation, selection, primer extension, and amplification, the positive control library member was enriched $>2.5 \times 10^7$ fold (Figure 4.7b), demonstrating the capability of this system to support highly effective iterative cycles of in vitro selection on modified nucleic acid polymers.

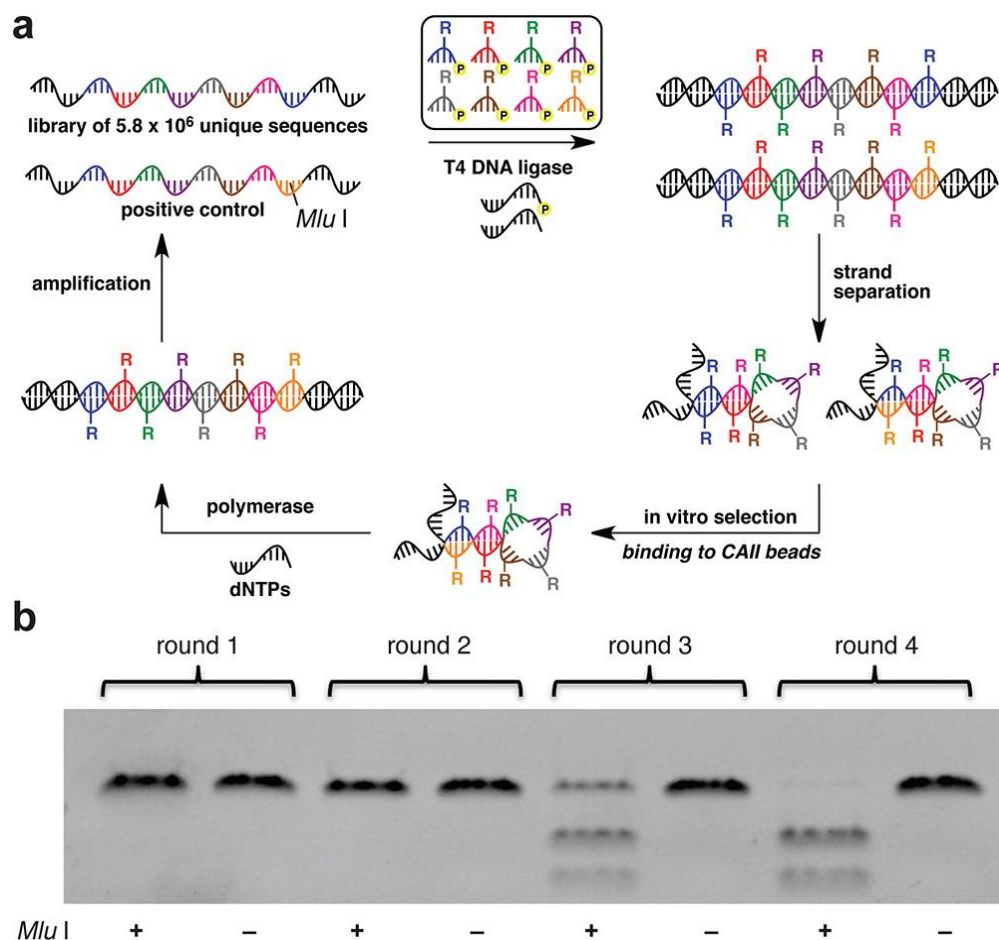


Figure 4.7. (a) Complete cycle of ligase-mediated translation, strand separation, selection, primer extension, and template amplification for functionalized nucleic acids. (b) PAGE analysis of *Mlu*I digestion products of a population of 5.8×10^6 DNA templates subjected to four iterated rounds of the cycle described in (a), resulting in the strong enrichment of a positive control template encoding a functionalized nucleic acid containing a carbonic anhydrase-binding pharmacophore.

4.8 Concluding Remarks

In Chapter Four, we have developed a strategy to translate DNA templates into sequence-defined highly functionalized nucleic acid polymers that uses T4 DNA ligase to catalyze the DNA-templated polymerization of functionalized trinucleotides. We incorporated

eight different functional groups throughout a polymer product, with the possibility of expanding the substrate set up to 64. In addition to exhibiting a high degree of sequence specificity, polymerization was remarkably efficient and could generate a polymer of 50 consecutive substrates (150 nucleotides), corresponding to a polymer of a molecular weight of approximately 60 kDa. The functionalized nucleic acid polymers were amenable to primer extension by Deep Vent (exo-) to regenerate the encoding template with high fidelity. Iterative cycles of translation, selection, template regeneration, and PCR amplification enabled the enrichment of a single library member encoding a carbonic anhydrase II inhibitor from a library of 5.8×10^6 highly functionalized DNAs.

Each of the two translation systems developed in this dissertation has their unique features. The enzyme-free translation system enables, for the first time, generation of sequence-defined non-nucleic acid synthetic polymers with backbone and side-chain structural diversity comparable to proteins. This novel category of synthetic polymers would allow exploration of unprecedented structure and sequence space through laboratory directed evolutions. The limitation of this system include lower polymerization yield compared to the enzymatic systems, and the difficult synthesis of macrocyclic substrates that restricts the scale of translation. The T4 DNA ligase-mediated translation system generates novel artificial polymers that are convenient to translate, select, synthesize, and develop, but their diversity may still be confined by their nature as nucleic acids. Developing together, two translation systems would introduce a wide array of novel structures within an evolvable synthetic polymer and therefore may help bridge the gap between biopolymers and synthetic polymers.

4.9 Methods

General Methods

Unless otherwise noted, all materials and compounds were prepared using commercially available reagents from Aldrich, and used without further purification. Water was purified with a Milli-Q purification system. DNA oligonucleotides shorter than 120 nucleotides were synthesized on a PerSeptive Biosystems Expedite 8909 DNA synthesizer; DNA oligonucleotides greater than 120 nucleotides were purchased from Integrated DNA Technologies and purified by PAGE. All materials and reagents used for oligonucleotide synthesis were purchased from Glen Research. All oligonucleotides were synthesized and deprotected according to the manufacturer's protocols. Oligonucleotides were purified by reverse-phase high-pressure liquid chromatography (HPLC, Agilent 1200) using a C18 stationary phase (Eclipse-XDB C18, 5 μ m, 9.4 x 200 mm) and an acetonitrile/ 100 mM triethylammonium acetate gradient. Oligonucleotide concentrations were quantitated by UV spectroscopy using a Nanodrop ND1000 spectrophotometer. Non-commercial oligonucleotides were characterized by LC/ESI-MS; reverse-phase separation was performed on an Alliance 2695 (Waters) HPLC system using a UPLC BEH C18 column (1.7 μ m, 2.1 x 50 mm) stationary phase and 6 mM aqueous triethylammonium bicarbonate/methanol mobile phase interfaced to a Q-Tof Micro mass spectrometer (Waters). Oligonucleotides greater than 70 nucleotides in length were analyzed by PAGE.

DNA Sequences

The sequences below are written from 5' \rightarrow 3'. = 3'biotinTEG (Glen Research, 20-2955); <P> = 5'Phosphate (Glen Research 10-1901); <Cam> = Amino-modifier C6 dC (Glen

Research 10-1019); <Gam> = N2-Amino-modifier C6 dG (Glen Research 10-1529); <Aam> = Aminomodifer C6 dA (Glen Research 10-1089); <Tam> = Amino-modifer C6 dT (Glen Research 10- 1039); <Cy5> = Cy5 phosphoramidite (Glen Research 10-5915); <5am5> = 5'-amino-Modifier 5 (Glen Research 10-1905); <IB> = iminobiotin (modified from <5am5> with EZ-link NHSiminobiotin (Thermo 21117), see amino-DNA modification below); <GLCBS> = Glycine- Leucine-4-carboxybenzenesulfonamide.¹

Templates

TA3: CCT GCC GTC GCA CTG CTG CTG CTG CTG CTG CTG CTG GAG CTG GCC
GCT

TB3: CCT GCC GTC GCA ACG ACG ACG ACG ACG ACG ACG ACG GAG CTG GCC
GCT

TC3: CCT GCC GTC GCA CGT CGT CGT CGT CGT CGT CGT CGT GAG CTG GCC
GCT

TD3: CCT GCC GTC GCA GCA GCA GCA GCA GCA GCA GCA GCA GAG CTG GCC
GCT

TE3: CCT GCC GTC GCA TGG TGG TGG TGG TGG TGG TGG TGG GAG CTG GCC
GCT

TF3: CCT GCC GTC GCA GAG GAG GAG GAG GAG GAG GAG GAG GAG CTG GCC
GCT

TG3: CCT GCC GTC GCA GCT GCT GCT GCT GCT GCT GCT GCT GAG CTG GCC
GCT

TH3: CCT GCC GTC GCA CGA CGA CGA CGA CGA CGA CGA GAG CTG GCC
GCT

TA-H: CCT GCC GTC GCA CTG ACG CGT GCA TGG GAG GCT CGA GAG CTG GCC
GCT

TA-HLP: CCT GCC CTC ACA GTC GCA CTG ACG CGT GCA TGG GAG GCT CGA GAG
CTG GGA TCT GCC GCT

TA-H1P: CCT GCC CTC ACA GTC GCA CTG ACG CGT GCA TGG GAG GCT CGA

T10C3: CCT GCC GTC GCA CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT GAG
CTG GCC GCT

T20C3: CCT GCC GTC GCA CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT
CGT CGT CGT CGT CGT CGT CGT CGT CGT GAG CTG GCC GCT

T30C3: CCT GCC GTC GCA CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT
CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT
CGT CGT GAG CTG GCC GCT

T40C3: CCT GCC GTC GCA CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT
CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT
CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT GAG CTG GCC GCT

T50C3: CCT GCC GTC GCA CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT
CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT
CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT CGT
CGT CGT CGT CGT CGT GAG CTG GCC GCT

TPos: CCT GCC CTC ACA GTC GCA CTG ACG CGT GCA TGG GAG GCT CGA GAG
CTG GGA TCT GCC GCT

Primers

FP1: <P>TGC GAC GGC AGG

FP2: AGC GGC CAG CTC

FP2cy5: <Cy5>AGC GGC CAG CTC

FLP1: <P> TGC GAC TGT GAG GGC AGG

FLP2: AGC GGC AGA TCC CAG CTC

RLP1: GAG CTG GGA TCT GCC GCT

RLP2: CCT GCC CTC ACA GTC GCA

RLP2: <5am5> CCT GCC CTC ACA GTC GCA

RLP2IB: <IB> CCT GCC CTC ACA GTC GCA

RLP2aT7: TAA TAC GAC TCA CTA TAG GGC TCG ATT TAA TTT CGC CGA CGT GAT

GAC ATT CCA GGC AGT CCT GCC CTC ACA GTC GCA

Amino-modified trinucleotides

A3-3: <P><Cam>AG

B3-3: <P><Cam>GT

C3-3: <P><Aam>CG

D3-3: <P><Tam>GC

E3-3: <P><Cam>CA

F3-3: <P><Cam>TC

G3-3: <P><Aam>GC

H3-3: <P><Tam>CG

C3-3glcbs: <P><GLCBS-Aam>CG

Terminator amino-modified trinucleotides

øA3-3: <Cam>AG

øB3-3: <Cam>GT

øC3-3: <Aam>CG

øD3-3: <Tam>GC

øE3-3: <Cam>CA

øF3-3: <Cam>TC

øG3-3: <Aam>GC

øH3-3: <Tam>CG

Synthesis of oligonucleotides

Synthesis of DNA templates and primers

DNA templates shorter than 120 nucleotides were synthesized on a PerSeptive Biosystems Expedite 8909 DNA synthesizer using a DMT-ON protocol on a 200 nmol scale (1000 Å CPG column). Following synthesis, the oligonucleotide was cleaved from the resin by incubation at 65 °C in 400 µL of a 1:1 mixture of ammonium hydroxide and methylamine for 15 minutes. The cleaved resin was filtered away by filtration, and the oligonucleotide was concentrated under reduced pressure using a speedvac. The residue was then taken up into 400 µL of 0.1M TEAA (pH 7), and purified using reverse-phase HPLC purification using a [8% acetonitrile in 0.1 M TEAA, pH 7] to [40% acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was then incubated at room temperature in 400 µL of 20% aqueous acetic acid for 1 h in order to cleave the DMT group. Following DMT deprotection, the oligonucleotide was frozen and lyophilized. The lyophilized product was taken up into 400 µL of 0.1 M TEAA and subjected to reverse-phase HPLC purification using a [8% acetonitrile in 0.1 M TEAA, pH 7] to [40% acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was then desalted by ethanol precipitation and dissolved into water at a 100 µM concentration.

Synthesis of 5'-phosphorylated primers

Primers were synthesized on a PerSeptive Biosystems Expedite 8909 DNA synthesizer using a DMT-ON protocol on a 200 nmol scale (1000 Å CPG column). Following synthesis, the oligonucleotide was cleaved from the resin by incubation at 65 °C in 400 µL of a 1:1 mixture of ammonium hydroxide and methylamine for 15 minutes. The cleaved resin was filtered away by filtration, and the oligonucleotide was concentrated under reduced pressure using a speedvac. The residue was then taken up into 400 µL of 0.1M TEAA (pH 7), and purified using reversephase HPLC purification using a [8% acetonitrile in 0.1 M TEAA, pH 7] to [40%

acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was then incubated at room temperature in 400 µL of 20% acetic acid for 1 h in order to cleave the DMT group. Following DMT deprotection, the oligonucleotide was frozen and lyophilized. The lyophilized product was then dissolved in 400 µL of ammonium hydroxide for 15 minutes at room temperature to remove the cyanoethyl modification, and then concentrated to dryness using a speedvac. The residue was then taken up into 400 µL of 0.1 M TEAA and subjected to reverse-phase HPLC purification using a [8% acetonitrile in 0.1 M TEAA, pH 7] to [40% acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was then desalted by ethanol precipitation and dissolved into water at a 100 µM concentration.

Synthesis of 5'-amino-modified primers

<5am5> modified DNA primers were synthesized on a PerSeptive Biosystems Expedite 8909 DNA synthesizer using a MMT-ON protocol on a 200 nmol scale (1000 Å CPG column). Following synthesis, the oligonucleotide was cleaved from the resin by incubation at 65 °C in 400 µL of a 1:1 mixture of ammonium hydroxide and methylamine for 15 minutes. The cleaved resin was filtered away by filtration, and the oligonucleotide was concentrated under reduced pressure using a speedvac; however, the solution was never evaporated to dryness (this can result in premature cleavage of the MMT protecting group, decreasing overall yields). The material was combined with 400 µL of 0.1 M TEAA (pH 7), and purified using reverse-phase HPLC using a [8% acetonitrile in 0.1 M TEAA, pH 7] to [40% acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was incubated at room temperature in 400 µL of 20% aqueous acetic acid for 1 h in order to cleave the MMT group. Following MMT deprotection, the oligonucleotide was frozen and lyophilized. The

lyophilized product was taken up into 400 μ L of 0.1 M TEAA and subjected to reverse-phase HPLC purification using a [8% acetonitrile in 0.1 M TEAA, pH 7] to [40% acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45 $^{\circ}$ C. The purified oligonucleotide was then desalted by ethanol precipitation and dissolved in water to a final concentration of 100 μ M.

Synthesis of 5'-iminobiotinylated primers for in vitro selection

To 120 μ L of a 40% aqueous DMSO solution was added: iminobiotin-NHS, 1.6 mM; 5'-modified primer, 85 mM; and TEA/HCl buffer (pH 10), 400 mM. The coupling reaction was stirred overnight at room temperature, and residual activated species were quenched by the addition of 50 ml 1 M Tris-Cl buffer (pH 8.0). The reaction was purified by reverse-phase HPLC using a [8% acetonitrile in 0.1 M TEAA, pH 7] to [40% acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45 $^{\circ}$ C. The purified oligonucleotide was desalted by ethanol precipitation and dissolved in water to a final concentration of 100 μ M.

Synthesis of functionalized trinucleotides

Synthesis of amino-modified trinucleotides

Trinucleotides were synthesized on a PerSeptive Biosystems Expedite 8909 DNA synthesizer using a DMT-ON protocol on a 1 μ mol scale (1000 \AA CPG column). Amino-modifier C6 dT (Glen, 10-1039), amino-modifier C6 dA (Glen, 10-1089), and amino-modifier C6 dC (Glen, 10-1019) were incorporated as specified by the manufacturer. Following synthesis, the oligonucleotide was cleaved from the resin by incubation at 65 $^{\circ}$ C in 400 μ L of a 1:1 mixture of ammonium hydroxide and methylamine for 15 minutes. The cleaved resin was removed by filtration, and the oligonucleotide was concentrated under reduced pressure using a speedvac.

The residue was then taken up into 400 μL of 0.1 M TEAA (pH 7), and purified using reverse phase HPLC purification using a [8% acetonitrile in 0.1 M TEAA, pH 7] to [40% acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was then incubated at room temperature in 400 μL of 20% acetic acid for 1 h in order to cleave the DMT group. Following DMT deprotection, the oligonucleotide was frozen and lyophilized. The lyophilized product was taken up into 400 μL of 0.1 M TEAA and subjected to reverse-phase HPLC purification using a [8% acetonitrile in 0.1 M TEAA, pH 7] to [40% acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was not desalted.

Functionalization of amino-modified trinucleotides with isothiocyanates

To 163 μL of DMSO was added 5'-phosphorylated amino-modified trinucleotide (25 nmol in 37 μL water), isothiocyanate (2.5 μmol in 25 μL of DMSO), and 25 μL of 1 M NaHCO_3 , and the reaction was incubated at room temperature for 12 h. The appearance of any precipitate did not influence the yield of reaction. The reaction was then quenched with 50 μL of 500 mM Tris-HCl (pH 8) and incubated at room temperature for 1 h. The mixture was then diluted to 500 μL with 0.1 M TEAA (pH 7) and then subjected to reverse-phase HPLC purification using a [8% acetonitrile in 0.1 M TEAA, pH 7] to [40% acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was not desalted.

Functionalization of amino-modified trinucleotides with cyclic anhydrides

To 163 μL of DMSO was added 5'-phosphorylated amino-modified trinucleotide (25 nmol in 37 μL water), cyclic anhydride (2.5 μmol in 25 μL of DMSO), and 25 μL of 1 M NaHCO_3 , and the reaction was incubated at room temperature for 12 h. The appearance of any

precipitate did not influence the yield of reaction. The reaction was then quenched with 50 μ L of 500 mM Tris-HCl (pH 8) and incubated at room temperature for 1 h. The mixture was then diluted to 500 μ L with 0.1 M TEAA (pH 7) and then subjected to reverse-phase HPLC purification using a [8% acetonitrile in 0.1 M TEAA, pH 7] to [40% acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was not desalted.

Functionalization of amino-modified trinucleotides with carboxylic acids

To 215 μ L of DMSO was added carboxylic acid (1.25 μ mol in 12.5 μ L DMSO), sulfo-*N*-Hydroxysuccinimide (sNHS) (3.33 μ mol in 10 μ L of 2:1 mixture of DMSO/H₂O), followed by 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) freebase (1.2 μ mol in 12 μ L of DMSO). The reaction was incubated at room temperature for 30 minutes, and then amino-modified trinucleotide (25 nmol in 15 μ L of water) was added, followed by 50 μ L of 500 mM NEt₃•HCl pH 10 buffer. The reaction was then incubated at room temperature for 12 hours and then diluted to 500 μ L using 0.1 M TEAA and subjected to reverse-phase HPLC purification using a [8% acetonitrile in 0.1 M TEAA, pH 7] to [40% acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was not desalted.

T4-DNA Ligase-mediate polymerization protocols

A. Polymerization on a hetero-octameric codon template

In a PCR tube was added DNA template (15 pmol in 1.5 μ L of water), initiating primer (60 pmol in 0.6 μ L of water), terminating primer (60 pmol in 0.6 μ L of water), functionalized trinucleotide library (60 pmol of each member in 5 μ L of water; 4 equivalents/codon), BSA (2 μ g in 1 μ L of water), 10 μ L of ligation buffer (132 mM Tris-HClHCl, 20 mM MgCl₂, 2 mM

dithiothreitol, 2 mM ATP, 15% Polyethylene glycol (PEG 6000), pH 7.6), 400 U of T4 DNA ligase (New England Biolabs, M0202), and water up to 20 μ L total volume. The polymerization was performed at 16 °C for 12 h and then desalted by gel filtration using CENTRI • SEP Spin Columns (Princeton Separations) equilibrated with water. The crude polymerized material was separated for analysis using non-denaturing PAGE (10% TBE, 150 V) and then stained by ethidium bromide and imaged by UV illumination and densitometry.

B. Strand separation of polymerized product

In a 1.5 mL centrifuge tube was added 30 μ L of Dynabeads MyOne Streptavidin C1 magnetic beads (Invitrogen, 650-01). The beads were washed (three washes using 200 μ L of binding buffer, 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM dithiothreitol, pH 7.9), and then the biotinylated polymerized product in 20 μ L of water was added along with 180 μ L of binding buffer. The suspension was incubated on a rotary at room temperature for 30 minutes. The beads were then washed three times with 200 μ L of binding buffer. 40 μ L of freshly prepared 150 mM NaOH was added to the beads, and the suspension was incubated at room temperature on a rotary for 15 minutes. The liberated strand was separated from the magnetic beads by use of a magnet, and the separated strand was immediately desalted by gel filtration using a CENTRI • SEP Spin Column (Princeton Separations) equilibrated with water.

C. Primer extension along functionalized polymer to regenerate a DNA template

To the crude single-stranded modified DNA was added 60 pmol of primer RPL2, 2 μ L of 10 X thermopol reaction buffer (NEB, B9004S), 2 μ L of 2 mM dNTPs mix, 2 U of Deep Vent

(exo-) and water up to 20 μ L. The mixture was incubated at 70 °C for 30 minutes and then the reaction was purified using a QIAQUICK nucleotide removal kit, eluting with water.

D. Sequence specificity assay by chain termination

To template TA-H1P (CCT GCC CTC ACA GTC GCA CTG ACG CGT GCA TGG GAG GCT CGA) (15 pmol) was added seven of the eight 5'-phosphorylated modified trinucleotides (60 pmol each from A3-3, B3-3, C3-3, D3-3, E3-3, F3-3, G3-3, H3-3) with the eighth modified trinucleotide being a terminator (60 pmol from \emptyset A3-3, \emptyset B3-3, \emptyset C3-3, \emptyset D3-3, \emptyset E3-3, \emptyset F3-3, \emptyset G3-3, \emptyset H3-3), initiating primer (60 pmol in 0.6 μ L of water), BSA (2 μ g in 1 μ L of water), 10 μ L of ligation buffer (132 mM Tris-HCl, 20 mM MgCl₂, 2 mM dithiothreitol, 2 mM ATP, 15% Polyethylene glycol (PEG 6000), pH 7.6), 400 U of T4 DNA ligase (New England Biolabs, M0202), and water up to 20 μ L total volume. The polymerization was performed at 16 °C for 12 h and then desalted by gel filtration using CENTRI • SEP Spin Columns (Princeton Separations) equilibrated with water. The crude polymerized material was separated for analysis using nondenaturing PAGE (10% TBE, 150 V) and then stained by ethidium bromide and imaged by UV illumination and densitometry. In eight separate experiments, the terminator was incorporated from the “A” position (first codon) to the “H” position (last codon).

In vitro selection protocol

STEP 1: In a PCR tube was added DNA library (1 pmol), initiating primer (4 pmol), terminating primer (4 pmol), functionalized trinucleotide library (10 pmol of each member), BSA (2 μ g in 1 μ L of water), 10 μ L of ligation buffer (132 mM Tris-HCl, 20 mM MgCl₂, 2 mM dithiothreitol, 2 mM ATP, 15% Polyethylene glycol (PEG 6000), pH 7.6), 400 U of T4 DNA

ligase (New England Biolabs, M0202), and water up to 20 μ L total volume. The polymerization was performed at 16 °C for 12 h and then desalted by gel filtration using CENTRI • SEP Spin Columns (Princeton Separations) equilibrated with water.

STEP 2: In a 1.5 mL centrifuge tube was added 30 μ L of Dynabeads MyOne Streptavidin C1 magnetic beads (Invitrogen, 650-01). The beads were washed (three washes using 200 μ L of binding buffer, 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM dithiothreitol, pH 7.9), and then the iminobiotinylated polymerized product in 20 μ L of water was added along with 180 μ L of binding buffer. The suspension was incubated on a rotary at room temperature for 30 minutes. The beads were then washed three times with 200 μ L of binding buffer. 40 μ L of freshly prepared 150 mM NaOH was added to the beads, and the suspension was incubated at room temperature on a rotary for 15 minutes. The liberated strand was separated from the magnetic beads by use of a magnet, and the separated strand was immediately desalted by gel filtration using a CENTRI • SEP Spin Column (Princeton Separations) equilibrated with water.

STEP 3: Bovine carbonic anhydrase II immobilized on carboxylic acid activated magnetic beads (MYONE Dynal Invitrogen) were freshly prepared according to a literature method using Carbonic anhydrase from bovine erythrocytes (Sigma-Aldrich, C3934).² The full volume of the positive control library of 5.8×10^6 unique sequences strand-separated material was incubated with 10 μ L of the carbonic anhydrase beads at room temperature for 30 minutes on a rotary. The beads were washed 10 times with PBS-Tween 20, 0.1% (PBS-T) 100 μ L, then the magnetic beads were resuspended in 50 μ L of distilled water and heated at 94 °C for 10 min. The beads were magnetically pelleted and the supernatant was recovered in a new tube and lyophilized.

STEP 4: To the enriched single-stranded modified DNA was added 4 pmol of primer RPL2, 2 μ L of 10 X thermopol reaction buffer (NEB, B9004S), 2 μ L of 2 mM dNTPs mix, 2 U of Deep Vent (exo-) and water up to 20 μ L. The mixture was incubated at 70 °C for 30 minutes and then the reaction was purified using a QIAQUICK nucleotide removal kit, eluting with water, diluted 100 fold and 1 μ L of this dilution was used for the PCR reaction using 1 μ M RLP2IB (<IB>CCT GCC CTC ACA GTC GCA), 1 μ M of FLP2 (AGC GGC AGA TCC CAG CTC), 1X IQ Supermix (Bio Rad, 170-8860), and water up to 400 μ L. The PCR was programed accordingly: 95 °C for 3 minutes, followed by 30 cycles of [95 °C for 30s, 55 °C for 30 s, 72 °C 30 s]. The crude PCR material was purified using QIAQUICK PCR purification kit (QIAGEN) eluting with water and lyophilized.

STEP 5: In a 1.5 mL centrifuge tube was added 30 μ L of Dynabeads MyOne Streptavidin C1 magnetic beads (Invitrogen, 650-01). The beads were washed (three washes using 200 μ L of binding buffer, 50 mM ammonium carbonate, pH 11, containing 0.5 M NaCl), and then the iminobiotinylated PCR product in 20 μ L of water was added along with 180 μ L of the binding buffer. The suspension was incubated on a rotary at room temperature for 30 minutes. The beads were then washed three times with 200 μ L of binding buffer. 40 μ L of freshly prepared 150 mM NaOH was added to the beads, and the suspension was incubated at room temperature on a rotary for 15 minutes. The liberated antisense strand was separated from the magnetic beads by use of a magnet. The basic wash process was repeated once more to ensure complete removal of the antisense strand. 40 μ L of 50 mM Tris-HCl, pH 6.8 containing 1 mM biotin was added to the beads and the eluted sense strand was desalted by gel filtration using a CENTRI • SEP Spin Column (Princeton Separations) equilibrated with water.

ENRICHMENT ANALYSIS: Following each round, 10 μ L of the crude PCR product was incubated at 37 °C for 1 h in 1X NEBuffer 3 with 10 U MluI (New England BioLabs, R0198). Following incubation, the restriction digest was directly analyzed on a 10% non-denaturing PAGE gel (150 V, 45 minutes) and visualized by ethidium bromide staining and UV illumination.

References

1. Lee, I. & Berdis, A. J. Non-natural nucleotides as probes for the mechanism and fidelity of DNA polymerases. *Biochim. Biophys. Acta* **1804**, 1064–80 (2010).
2. Obeid, S., Baccaro, A., Welte, W., Diederichs, K. & Marx, A. Structural basis for the synthesis of nucleobase modified DNA by *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 21327–31 (2010).
3. Henry, A. & Romesberg, F. E. The evolution of DNA polymerases with novel activities. *Curr. Opin. Biotechnol.* **16**, 370–7 (2005).
4. Ghadessy, F. J., Ong, J. L. & Holliger, P. Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4552–7 (2001).
5. Kranaster, R. & Marx, A. Engineered DNA polymerases in biotechnology. *Chembiochem* **11**, 2077–84 (2010).
6. Ong, J. L., Loakes, D., Jaroslowski, S., Too, K. & Holliger, P. Directed evolution of DNA polymerase, RNA polymerase and reverse transcriptase activity in a single polypeptide. *J. Mol. Biol.* **361**, 537–50 (2006).
7. Loakes, D., Gallego, J., Pinheiro, V. B., Kool, E. T. & Holliger, P. Evolving a polymerase for hydrophobic base analogues. *J. Am. Chem. Soc.* **131**, 14827–37 (2009).
8. Ramsay, N. *et al.* CyDNA: synthesis and replication of highly Cy-dye substituted DNA by an evolved polymerase. *J. Am. Chem. Soc.* **132**, 5096–104 (2010).
9. Loakes, D. & Holliger, P. Polymerase engineering: towards the encoded synthesis of unnatural biopolymers. *Chem. Commun.* 4619–31 (2009). doi:10.1039/b903307f
10. Chaput, J. C. & Szostak, J. W. TNA synthesis by DNA polymerases. *J. Am. Chem. Soc.* **125**, 9274–5 (2003).
11. Leconte, A. M. *et al.* Directed Evolution of DNA Polymerases for Next-Generation Sequencing. *Angew. Chem. Int. Ed.* 1–5 (2010). doi:10.1002/anie.201001607
12. Pinheiro, V. B. *et al.* Synthetic Genetic Polymers Capable of Heredity and Evolution. *Science* **336**, 341–344 (2012).
13. Malyshev, D. a *et al.* Efficient and sequence-independent replication of DNA containing a third base pair establishes a functional six-letter genetic alphabet. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 12005–10 (2012).

14. Jäger, S. & Famulok, M. Generation and enzymatic amplification of high-density functionalized DNA double strands. *Angew. Chem. Int. Ed.* **43**, 3337–40 (2004).
15. Nair, P. a *et al.* Structural basis for nick recognition by a minimal pluripotent DNA ligase. *Nat. Struct. Mol. Biol.* **14**, 770–8 (2007).
16. Odell, M., Sriskanda, V., Shuman, S. & Nikolov, D. B. Crystal Structure of Eukaryotic DNA Ligase – Adenylate Illuminates the Mechanism of Nick Sensing and Strand Joining. **6**, 1183–1193 (2000).
17. Raae, A. J., Kleppe, R. K. & Kleppe, K. Kinetics and effect of salts and polyamines on T4 polynucleotide ligase. *Fed. Eur. Biochem. Soc. J.* **60**, 437–443 (1975).
18. Yuan, C., Lou, X. W., Rhoades, E., Chen, H. & Archer, L. A. T4 DNA ligase is more than an effective trap of cyclized dsDNA. *Nucleic Acids Res.* **35**, 5294–5302 (2007).
19. Nilsson, S. V & Magnusson, G. Sealing of gaps in duplex DNA by T4 DNA ligase. *Nucleic Acids Res.* **10**, 1425–1437 (1982).
20. Kim, H., Lee, H. J. & Pack, M. Characteristic Mode of Action of T4 DNA Ligase. *jbmborkr* **19**, 301~306 (1986).
21. Patel, M. P., Baum, D. A. & Silverman, S. K. Improvement of DNA adenylation using T4 DNA ligase with a template strand and a strategically mismatched acceptor strand. *Bioorg. Chem.* **36**, 46–56 (2008).
22. Kotler, L. E., Zevin-Sonkin, D., Sobolev, I. a, Beskin, a D. & Ulanovsky, L. E. DNA sequencing: modular primers assembled from a library of hexamers or pentamers. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 4241–5 (1993).
23. Kaczorowski, T. & Szybalski, W. Genomic DNA sequencing by SPEL-6 primer walking using hexamer ligation. *Gene* **223**, 83–91 (1998).
24. Mendel-Hartvig, M., Kumar, A. & Landegren, U. Ligase-mediated construction of branched DNA strands: a novel DNA joining activity catalyzed by T4 DNA ligase. *Nucleic Acids Res.* **32**, e2 (2004).
25. Liang, X., Fujioka, K. & Asanuma, H. Nick sealing by T4 DNA ligase on a modified DNA template: tethering a functional molecule on D-threoninol. *Chem. Eur. J.* **17**, 10388–96 (2011).
26. Kaczorowski, T. & Szybalski, W. Co-operativity of hexamer ligation. *Gene* **179**, 189–93 (1996).

27. Yoshino, Y., Ishida, M. & Horii, A. A new 10-min ligation method using a modified buffer system with a very low amount of T4 DNA ligase: the “Coffee Break Ligation” technique. *Biotechnol. Lett.* **29**, 1557–60 (2007).
28. Cho, G., Keefe, a D., Liu, R., Wilson, D. S. & Szostak, J. W. Constructing high complexity synthetic libraries of long ORFs using in vitro selection. *J. Mol. Biol.* **297**, 309–19 (2000).
29. Hirao, I., Kimoto, M. & Yamashige, R. Natural versus Artificial Creation of Base Pairs in DNA: Origin of Nucleobases from the Perspectives of Unnatural Base Pair Studies. *Acc. Chem. Res.* **45**, 2055–65 (2012).
30. McGregor, L. M., Gorin, D. J., Dumelin, C. E. & Liu, D. R. Interaction-dependent PCR: identification of ligand-target pairs from libraries of ligands and libraries of targets in a single solution-phase experiment. *J. Am. Chem. Soc.* **132**, 15522–4 (2010).